

Empirical Bayes, Bayes factors and deoxyribonucleic acid fingerprinting

Ruma Basu

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2017

ABSTRACT

Empirical Bayes, Bayes factors and deoxyribonucleic acid fingerprinting

Ruma Basu

The central theme in this thesis is Empirical Bayes. It starts off with application of Bayes and Empirical Bayes methods to deoxyribonucleic acid fingerprinting. Different Bayes factors are obtained and an alternative Bayes factor using the method of Savage is studied both for normal and non normal priors. It then moves on to deeper methodological aspects of Empirical Bayes theory. A 1983 conjecture by Carl Morris on the parametric empirical Bayes prediction intervals for the normal regression model is studied and an improvement suggested. Carlin and Louis' (1996) parametric empirical Bayes prediction interval for the same model is also dealt with analytically while their approach had been primarily numerical. It is seen that both of these intervals have the same coverage probability up to a certain order of approximation and they have the same expected length up to the same order of approximation. Both the intervals are equal tailed up to the same order of approximation. Then the corrected proof of an important published result by Datta, Ghosh and Mukerjee (2000) is provided using first principles of probability matching. This result is relevant to our work on parametric empirical Bayes prediction intervals.

TABLE OF CONTENTS

CHAPTER		PAGE
1	A Brief Introduction	
1.	Introduction	1
2.	A survey of some important work in empirical Bayes and a review of Chapter 3	1
3.	A brief history of the DNA fingerprinting problem and a review of Chapter 2	5
4.	Probability matching priors and a review of Chapter 4	7
2	A Bayes and empirical Bayes approach to DNA fingerprinting	
1.	Introduction	9
2.	The method of Match Binning	9
3.	The Bayesian formulation as in Berry	12
4.	Parametric empirical Bayes formulation of the DNA fingerprinting problem	17
5.	The alternative Bayes factor	20
3	Settling a conjecture of Morris	
1.	Introduction	26
2.	Morris's interval and modification thereof	27
3.	Another interval	28
4.	Expected lengths	30
4	Rigorous justification of the prior of Datta et al. (2000)	
1.	Introduction	31
2.	Review and main result of Datta et al. (2000)	31
3.	Modified derivation of the result in section 3 of Datta, Ghosh and Mukerjee (2000)	33
	REFERENCES	40
	APPENDIX	47

Acknowledgements

I take this opportunity to thank all those persons without whose support and help the thesis would not have been completed. My sincere thanks to Prof. Shaw Hwa Lo who gave me the first problem, the starting point of the thesis and worked hard to allow me to submit at Columbia University even after such a long gap. Without his efforts all my work would have gone in vain.

Prof. Jayanta Kumar Ghosh guided me during my long and difficult years in India and without his advice, guidance and inspiration this thesis would not have happened. My gratitude towards Prof. Ghosh knows no end. Mere words are not enough to express what he had done for me. I thank him from the bottom of my heart.

Prof. Rahul Mukerjee collaborated with Prof. Ghosh and me on the material covered in Chapters 3 and 4 and without his help this work might not have been possible. The material in Chapter 3 got published as a joint paper and they have kindly allowed me to present the material in my thesis. I convey my sincere thanks to Prof. Mukerjee.

I am grateful to the Chairman and faculty members of the Department of Statistics, Columbia University for all their assistance. I must mention the name of Mrs Dood Kallicharan who has been a good friend to me throughout and was always ready to help me.

I am grateful to the Indian Statistical Institute, Kolkata for allowing me to use many of its facilities at different points of time. The faculty members of the Institute have been immensely supportive and I benefitted from many useful discussions with them. Special mention needs to be made of Prof. Partha P. Majumdar and Prof. Tapas Kumar Samanta. I have to mention my friends and family who were a constant source of support and encouragement during my difficult years. I wish I could name them all but the list would go on. To each of them I owe a lot. Dr. Jyotishka Datta and Dr. Kajal Dihidar helped me with programming in R. I am in their debt. I convey my thanks to Mr. Dibyendu Bose for meticulously typing the thesis in Latex and to Dr. Swagata Nandi and Dr. Anirvan Chakraborty for answering many Latex related queries.

To
Ma and Baba

Chapter 1

A Brief Introduction

1.1 Introduction

This work began as a parametric empirical Bayes approach to DNA fingerprinting, which was then a novel technology as well as a novel methodology. But very soon it went deep into questions of testing and model selection in a much more general framework. From DNA fingerprinting and methodology of testing and model selection, I moved to two major problems of parametric empirical Bayes methodology which required deep technicalities. This chapter provides an overview of both problems discussed and results obtained in the thesis. Each section below deals with one later chapter, but not quite in the order in which the chapters appear, as briefly explained below.

Empirical Bayes methods have been central to the thesis. Thus instead of starting with a review of Chapter 2, the DNA fingerprinting problem, I begin with a survey of some important work in empirical Bayes and a review of our Chapter 3 where we work out a modified version of Morris' (1983) parametric empirical Bayes prediction intervals using a normal regression model and higher order asymptotics. For the same model we work on the idea of Carlin and Louis (1996) to obtain a second interval. Section 1.3 covers a history of the DNA fingerprinting problem and the statistical methodologies developed and reviewed by us in Chapter 2. In Section 1.4 we provide an overview of probability matching and discuss a corrected proof of an important published result by Datta, Ghosh and Mukerjee(2000).

1.2 A survey of some important work in empirical Bayes and a review of Chapter 3.

The journey of empirical Bayes started with the path breaking paper by H.E. Robbins (1956) and it gained momentum and support with the paper of Neyman (1962). Neyman (1962) described Robbins's work as a breakthrough. The name empirical Bayes was coined by Robbins in his 1956 paper and it seemed like an 'oxymoron' in the beginning. What Robbins did came to be known as General empirical Bayes or Non parametric empirical Bayes. The 1956 paper dealt with n structurally similar problems of estimating

θ_i from X_i , ($i = 1, 2, \dots, n$) where X_i has density $p_{\theta_i}(x_i)$ and the independent θ_i 's can be assumed to have the same prior distribution π . $X_i|\theta_i$ are independent and the form of the prior density is unknown but assumed to be existent. It is known that with squared error loss function the Bayes point estimator for θ_i is $E(\theta_i|X_i)$ which cannot be calculated though without a known form of the prior. Robbins worked with the following discrete distributions-Poisson, Geometric and the ‘‘Laplacian’’ and suggested estimating the marginal densities on the basis of X_1, X_2, \dots, X_n and replacing them in the expression for $E(\theta|X)$, leading to the empirical Bayes estimator.

The poisson distribution, for instance, gives $E(\theta|x)$ to be $(x+1)f(x+1)/f(x)$ which he estimated by $(x+1)(\# \text{ of observations} = x+1)/(\# \text{ of observations} = x)$.

Thus we see that inference about a certain θ involves terms related to the marginal density at its observed value and the marginal density at $x+1$, leading to borrowing of information from its neighbours. Similar concepts appear in Gini (1911) and mentioned by both Forcina (1982, pg. 68) and Robbins (1983).

Robbins followed up the 1956 paper with papers on linear empirical Bayes which impose certain conditions on the mean and variance (1983). Post 1956 one saw a burgeoning of literature in nonparametric empirical Bayes by students and descendants of Robbins, and their students and so on.

Before we move on to Parametric empirical Bayes let us discuss 2 papers by Stein (1956) and James and Stein (1961) which belong to the Compound theory in the frequentist domain. Stein(1956) showed that the usual maximum likelihood estimator, \tilde{X} , for the mean of a multivariate normal distribution was inadmissible. James and Stein(1961) showed, for $k \geq 3$, the exact form of the estimator, which was doing better than the usual ML estimator with squared error loss and the estimate for the i th parameter value depended on all the other observations. This was a shocking revelation at the time although this was also not admissible as they obtained the positive part James-Stein estimator which was doing even better than the original James-Stein estimator.

Stein and James-Stein’s work were put into an empirical Bayes context in a series of papers by Efron and Morris (1971, 1972, 1973, 1975) and Morris (1983). Henceforth, the work will come under the purview of Parametric empirical Bayes where a parametric form for the prior is chosen but the prior parameters are unknown and the prior parameters are estimated from the data using the method of Maximum Likelihood II prior estimation. In MLII prior estimation one takes the likelihood function to be the marginal density as a function of the parameter(s) and maximises that with respect to the parameters thus obtaining the type II maximum likelihood prior or MLII prior in short. In complicated

situations one takes recourse to the *E-M* algorithm of Dempster, Laird and Rubin (1977) but we do not discuss the method here.

Efron and Morris (1973) showed that the James-Stein estimator can be obtained as a parametric empirical Bayes estimator with the following model : X_1, X_2, \dots, X_k follows the normal $(\theta_i, 1)$ distribution and are independent. θ_i follows the normal $(0, A)$ distribution and are independent, A is unknown and is estimated from the data leading to the James-Stein estimator for $k \geq 3$.

In Morris (1983), this theory is shown to provide a framework for PEB inference for a broad class of very interesting practical problems. For example it was shown in Morris (1983) that the empirical Bayes predictive interval for baseball players, based on data from the first half of the season was much shorter than usual confidence intervals. Morris has continued to do such applied work based on the EB methodology, a major application being in the ranking of hospitals.

Morris (1983) deals with a normal regression model where the observations $Y_i|\theta_i$ follow the normal (θ_i, V) distribution and are independent, $\theta_i|\beta, A$ follows $N(z'_i\beta, A)$ and are independent. $i = 1, 2, \dots, k$, and z_i 's are the r dimensional column vectors of regression variables and β a vector of regression coefficients. $0 \leq r \leq k - 3$.

Morris works on both the equal variance case and unequal variances case but we shall discuss in brief only the equal variance case as our work on parametric empirical Bayes prediction intervals deals only with the equal variance case. The common variance V and the z_i 's are assumed to be known. It is required to estimate the unknown θ_i 's based on the observed Y_i 's.

The posterior distribution of $\theta_i|Y_i, \beta, A$ is obtained as Normal, the posterior mean being $((1-B)Y_i + Bz'_i\beta)$, posterior variance $V(1-B)$ where $B = V/(V+A)$. With squared error loss the estimated posterior mean is taken as the empirical Bayes estimate for θ_i , the estimates for β and B being obtained via complete sufficient unbiased estimators (or *MVUEs*) of β and B .

$\hat{\beta} = b = (z'z)^{-1}z' Y$ and $\hat{B} = (k - r - 2)V/S$ where $S = \sum(Y_i - z'_ib)^2$

Thus the empirical Bayes estimator for θ_i equals $(1 - \hat{B})Y_i + \hat{B}z'_ib$

This estimator, called the compromise estimator by Morris, is the celebrated James-Stein estimator also. Stein further improved this estimate by imposing $\hat{B} \leq 1$ and taking \hat{B} to be equal to

$$((k - r - 2)/(k - r))V/(V + \hat{A}^+) \text{ where } \hat{A} = S/(k - r) - V \text{ and } \hat{A}^+ = \max(0, \hat{A}) \quad (*)$$

Following Morris we define the parametric empirical Bayes prediction interval for θ_i . Let $\xi = (\beta', A)'$. Then a PEB prediction interval for θ_i of size $(1 - \alpha)$ is a random set C , depending only on Y_1, Y_2, \dots, Y_k and satisfying the conservative PEB requirement $P_\xi(\theta_i \text{ lies in } C) \geq 1 - \alpha$ for all ξ .

In our work we have defined the following :

Z_k is the $k \times r$ design matrix with rows z'_1, z'_2, \dots, z'_k . As in Morris we assume $r \leq k - 3$ and that Z_k has full column rank. The following assumption is also needed. The sequence $c_{ik} = k\{z'_i(Z'_k Z_k)^{-1} z_i\}$ is bounded.

$$\begin{aligned}
b &= (Z'_k Z_k)^{-1} Z'_k Y \\
s^2 &= (k - r)^{-1} \sum (Y_j - z'_j b)^2 \\
B &= V/(V + A) \\
\hat{B} &= ((k - r - 2)/(k - r))V/\max(V, s^2) \\
\rho &= B/(1 - B), \\
\hat{\rho} &= \hat{B}/(1 - \hat{B}) \\
h_{ik}(u; \rho) &= \frac{1}{2}(c_{ik} - r)\rho + \left(\frac{1}{4}\right)(u^2 + 1)\rho^2 \quad (-\infty < u < \infty) \\
\hat{\theta}_i &= (1 - \hat{B})Y_i + \hat{B}z'_i b \\
s_i^2 &= V(1 - ((k - r)/k)\hat{B}) + \{2/(k - r - 2)\}\hat{B}^2(Y_i - z'_i b)^2
\end{aligned}$$

It is known from standard asymptotics that the naive interval given by

$$(\hat{\theta}_i - z_{\alpha/2} \sqrt{V(1 - \hat{B})}, \hat{\theta}_i + z_{\alpha/2} \sqrt{V(1 - \hat{B})})$$

is an approximate empirical Bayes prediction interval and is of order $O(1/k)$. Morris points out that the above does not take into consideration the fact that B and β are estimated from the data. Morris suggests a bigger interval given by

$$(\hat{\theta}_i - z_{\alpha/2} s_i, \hat{\theta}_i + z_{\alpha/2} s_i)$$

and claims that its coverage probability is at least $1 - \alpha$. In chapter 3 we provide an example where this does not hold true and go on to suggest two parametric empirical Bayes prediction intervals, for the above model, which are $o(1/k)$. The first interval given by $I = [\hat{\theta}_i \mp z\{1 + k^{-1}h_{ik}(z; \hat{\rho})\}s_i]$ is a modification of Morris's interval. Theorem 2.1 of Chapter 3 is crucial in proving the coverage probability of the above interval. And from the same theorem we can see that the interval is equal tailed up to the same order of approximation.

The second interval is obtained by exploring analytically Carlin and Louis' (1996) idea. Let us define

$$\begin{aligned} B^* &= V/\max(V, (k-r)s^2/k), \text{ the MLE of } B \\ \theta_i^* &= (1-B^*)Y_i + B^*z_i'b \\ h_{ik}^*(u; \rho) &= \frac{1}{2}(c_{ik} + r + 4)\rho + \left(\frac{1}{4}\right)(u^2 + 1)\rho^2 \end{aligned}$$

$\rho^* = B^*/(1-B^*)$ if $B^* < 1$ and defined arbitrarily if $B^* = 1$. Then the second interval is given by

$$I^* = [\theta_i^* \mp z\{1 + k^{-1}h_{ik}^*(z; \rho^*)\}\{V(1-B^*)\}^{1/2}]$$

which also has the same coverage probability upto $o(1/k)$ and is equal tailed upto $o(1/k)$. Theorem 3.1 of Chapter 3 helps us obtain the coverage probability but its proof is omitted as it is similar to theorem 2.1.

Theorem 4.1 shows that the above 2 intervals have the same expected length also up to the same order of approximation.

1.3 A brief history of the DNA Fingerprinting problem and a review of Chapter 2:

DNA Fingerprinting was accidentally invented by Sir Alec Jeffreys at the University of Leicester in 1984. He was looking at the X-Ray film of DNA samples of different members of his technician's family and he got to notice some similarities and dissimilarities and the basic idea of DNA Fingerprinting dawned on him. Around the same time the Polymerase Chain Reaction (PCR) technique for making copies of DNA samples was being perfected by Kary Mullis in the USA. Since then DNA fingerprinting coupled with the PCR has been used to solve many criminal cases involving rape and murder, and immigration as well as paternity cases. The first such case to be solved by Jeffreys was an immigration case of a British boy whose family was from Ghana. The DNA Fingerprinting test showed similarities in the DNA of his family members and that of the boy thus inferring that he belonged to that family. The next important case was the rape case of two teenagers in 1983 and 1986 where the prime suspect was exonerated while the real culprit was convicted based on DNA samples obtained from semen from the two victims. Thus started the journey of DNA Fingerprinting and has spread worldwide since then.

Points on the DNA which are highly variable are usually measured and compared and

these points are called hypervariable loci. In the 1990's work started on micro-satellite regions on the DNA called Short Tandem Repeat or STR. Peter Gill of the Forensic Science Service and his team of Scientists perfected the method of STR profiling leading to the UK National DNA data base (NDNAD) in 1995, where observations on the DNA samples of all those arrested in UK are collected and stored for future use. A similar database called CODIS has been developed in the USA based on 13 specific STR loci, chosen by the FBI. The likelihood that any 2 individuals (except identical twins) will have the same 13 loci DNA profile can be as high as 1 in billion or greater.

Our work on DNA Fingerprinting was inspired by Don Berry (1991) and we also look at some Bayesian methods developed by Lindley (1977) in a more general context of Identification problems and refined by Berry. In the early 1990's DNA Fingerprinting had become a hot subject and began to be used in courts. This led to various statistical issues. Should one analyse the data as a classical statistician with P -values or the method of match binning or should one be a Bayesian and try to calculate or estimate a posterior probability of guilt of the suspect? Berry and several other statisticians like Evett, Pinchin, Roeder, Geisser, Devlin and Risch among others explored several questions relating to these and we mention those in brief below. In England and France the Bayesian method became popular whereas the method of match binning was being used in the US. Our work on the use of parametric empirical Bayes methods in the case of DNA fingerprinting led us to our work on methodological aspects of parametric empirical Bayes in Chapters 3 and 4.

In Chapter 2 we first review the classical method of match binning and the method of evaluating Bayes factors based on Berry's work. In the Bayesian paradigm choice of prior becomes an important issue. Berry first works with a simple model where the prior is assumed to be improper uniform on $(-\infty, \infty)$ leading to a nice attractive formula. Next he works on the more realistic model where he estimates the prior in a frequentist manner based on the reference population. A lot of questions have been raised about his method of density estimation which we discuss in section 2.3.

In section 2.4 the parametric empirical Bayes formulation of the problem is made using normal model for the data and the prior, the prior parameters being unknown and estimated by the method of ML II prior estimation discussed in section 1.2.

In section 2.5 we review Lindley's (1977) work and proceed to suggest an alternative Bayes factor for the same model. The prior under the null in the case of Bayes factor enumeration is usually done by the method of H. Jeffreys. We calculate the Bayes factor using Savage's method of enumerating the prior under the null by proper conditioning.

This helps us to remove a term in the Bayes factor which has a confusing presence. The alternative Bayes factor is also calculated for a general non-normal prior and we show how this can be calculated without sophisticated density estimation methods that have been discussed in section 2.3.

In this context it appears to us that it is still not clear which of the two Bayes Factors is appropriate in a particular problem. Further research on this is desirable.

1.4 Probability matching priors and a review of Chapter 4 :

A prior is probability matching if it is such that the posterior credibility interval based on the prior is approximately the same as the frequentist confidence interval based on the maximum likelihood estimator up to a certain order of approximation. Essentially it is matching two kinds of probabilities - the classical and the posterior. The credibility interval based on this prior can be interpreted as a frequentist or classical confidence interval.

A probability matching prior or PMP is called a first order PMP when the order of approximation is $o(n^{-1/2})$ where n is the sample size. Mukerjee and Dey's (1993) work narrowed down the class of first order PMP by imposing the condition that the order of approximation should be $o(n^{-1})$ thus leading to second order PMPs.

The matching criterion leads to a partial differential equation for the PMP depending on the order of approximation. Solving the partial differential equation one arrives at the PMP.

Thus probability matching helps one to obtain a non subjective Bayesian prior and inference based on a PMP does not give rise to something that is purely subjective. In this sense the prior, which is found in a purely algorithmic way, may be regarded as non-subjective, i.e., the prior has no subjective input. Objective Bayesians do not distinguish between non-subjective and objective priors, so a probability matching prior may be considered to be an objective prior.

First order probability matching based on posterior quantiles was first investigated by Welch and Peers (1963). He worked with a scalar parameter θ thus not considering any nuisance parameters and obtained Jeffrey's prior as the first order PMP. Peers (1965), Stein (1985), Tibshirani (1989), Nicolau (1993), Datta and Ghosh (1995a,b), Ghosh, Carlin and Srivastava (1995), Ghosh and Yang (1996) and Garvan and Ghosh (1997) among others pursued the idea of Welch and Peers to work on examples with nuisance parameters. Several authors have noted that there are infinitely many first order probability

matching priors in the presence of nuisance parameters.

Work on second order PMPs was pursued by Sun and Ye (1996), Mukerjee and Ghosh (1997) and Garvan and Ghosh (1997) among others. Brief reviews of PMPs can be obtained in Ghosh and Mukerjee (1997), Ghosh and Mukerjee (1992a), Reid (1995) and Kass and Wasserman (1996).

Datta, Ghosh and Mukerjee (2000) worked on a normal hierarchical Bayes model and showed that the probability matching prior obtained is different from Morris's uniform priors for the mean and variance and the exact form of the prior for the variance is obtained. But the proof is incorrect - the reasons being mentioned in Section 4.2. We provide an alternative proof of the above result which relies on first principles and is obtained starting from the definition of the matching criterion. We see that the result obtained in Datta et al. (2000) holds true irrespective of the problem in the proof.

For the sake of completeness we mention the model here. Given $\theta_1, \theta_2, \dots, \theta_n$ the random variables Y_1, Y_2, \dots, Y_n are independent and Y_i follows Normal $(\theta_i, V), i = 1, 2, \dots, n$. Also $\theta_1, \theta_2, \dots, \theta_n$ are *i.i.d* such that for each i θ_i is Normal with mean ξ_1 and variance $\xi_2(-\infty < \xi_1 < \infty, \xi_2 > 0)$. V is assumed to be known while ξ_1, ξ_2 are unknown parameters, the distributions of which are obtained through probability matching. They assume the prior to be of the form $\pi(\xi_1, \xi_2) = \text{constant} \cdot \pi(\xi_2)$. We work with the same form and obtain exactly the results obtained in Datta *et al.*, that is, the prior for the mean is the uniform on $(-\infty, \infty)$ and is independent of the prior for the variance which is proportional to $\xi_2(V + \xi_2)^{-1}$. Thus we see that the PMP's for the above model is slightly different from the Laplacian priors used by Morris (1981) and hence the confidence intervals mentioned by Morris (1983) cannot be expected to be of the order $o(n^{-1})$.

The Shrinkage Argument or a Bayesian route to frequentist calculations has been used in section 4.3 and so we discuss it in brief here. The method was first suggested by J.K.Ghosh and its early applications can be found in Bickel and Ghosh (1990), Dawid (1991), Ghosh and Mukerjee (1991) and Ghosh (1994, Chapter 9). The method is presented in detail in Mukerjee and Reid (2000) and also in section 1.2 of the monograph on PMPs by Datta and Mukerjee (2004). The name is derived from the fact that a sequence of priors is chosen in the procedure and it is allowed to converge weakly to a degenerate prior. This method significantly reduces the algebra underlying frequentist higher order asymptotic computations.

Chapter 2

A Bayes and Empirical Bayes approach to DNA Fingerprinting

2.1 Introduction

DNA Fingerprinting has become a major new source of evidence in legal cases, for both the defence and the prosecution. Very briefly the basic statistical issues may be formulated as follows :

A crime has been committed. Some biometric evidence has been left at the crime scene by the criminal. We denote this by a random piece of data Y which can be discrete or continuous. In problems of DNA Fingerprinting Y is typically a small part or small parts of the DNA recovered from the body fluids left at the scene of the crime. Typically, the part of the DNA used as evidence shows a lot of polymorphic variation in the relevant population of possible suspects. The police have a suspect from whom a similar biometric evidence, X has been collected. The basic question is whether $X = Y$ (at least approximately) in which case there is a match. If $X \neq Y$ (even under the above approximate sense) there is no match.

A match suggests the suspect may be guilty. Absence of a match suggests the suspect is innocent. Under the usual notions of justice and evidence (in Britain, U.S.A., India etc.) the suspect is innocent unless the weight of evidence against the suspect is strong. The evidence against the suspect is considered to be strong if there is a match and the match is improbable if one chose at random from the population of suspects.

How do we make these things precise ? The answer depends on probabilistic calculations. Statistics has now two fundamental paradigms, Classical Statistics and Bayesian Analysis. The calculation of probability depends on the paradigm. We first deal with the classical method in the next section. The entire chapter is joint work with Prof. J.K. Ghosh.

2.2 The method of Match Binning

X, Y mentioned in the previous section and all other measurements on the weights of DNA pieces come with an error in measurement which is small and is proportional to

the true band weights. Let μ and ν be the true band weights of X and Y respectively. Distributional assumptions are made on X and Y and both the Normal and Lognormal distributions seem to give a good fit. Berry (1991) works on the log scale, we have mostly used the Normal distribution. (Distributional assumptions are not required for the method of Match Binning).

Another point to note is that the DNA fragment consists of two strands, one from the father and one from the mother and usually it is not known which is which. So at each site we have two figures giving two weights.

Readings can come from a single site (single locus probe) or from multiple sites (multiple locus probe). We have confined our attention to single locus probes.

The basic question whether X and Y came from the same person or not can be formulated as a problem of hypothesis testing, where the null hypothesis is Innocence ($\mu \neq \nu$) and the alternative is Guilt ($\mu = \nu$). Thus we have

To Test $H_0 : \mu \neq \nu$ (Innocence)

versus $H_1 : \mu = \nu$ (May be Guilty)

since the status quo is Innocence and one is called guilty only if there is compelling evidence to the contrary. Standard methods of testing do not work for this non-standard hypothesis testing problem.

Let us note that a null hypothesis like this also occurs in problems of Bio-equivalence where it is claimed that a new drug (with mean ν for the variable under consideration) has same mean as the old drug (with mean μ). Null hypotheses of this kind are not easy to handle in classical statistics. Usually, in problems of Bio-equivalence a suitable threshold $\delta > 0$ is chosen and H_0 is taken to be $|\mu - \nu| \geq \delta$. Under certain conditions it is known that among tests with probability of error of first kind $\leq \alpha$, there is a *UMP* test for μ, ν with $H_1 : |\mu - \nu| < \delta$, vide Lehmann (1986).

The problem of Bio-equivalence is actually much simpler than the problem of establishing Guilt as far as Classical Statistics is concerned. When we try to establish Guilt we use as known a database of random sample from a population to which the suspect is supposed to belong. It is difficult to incorporate such knowledge in Classical Statistics. The method of Match-Binning had been developed to deal with this problem.

The standard deviation, c , of the error in measurement is assumed to be known. It is estimated from previous experimental data and is generally of the order of .0042.

A match is said to occur if $|x - y| < Kc\sqrt{2}$, K being fixed arbitrarily depending on the stringency requirement in the test and it usually varies between $K = 2/3$ to $K = 3$. A small value of K indicates that there is a large proportion of guilty individuals being

set free and receiving the benefit of doubt while a large K involves a higher risk of false conviction for innocent individuals. To protect innocent individuals we advocate using a small value of K .

H_1 is rejected outright if there is no match. In case there is a match one goes on to calculate the match proportion. For this a database is required, of similar observations, from a population of individuals, referred to as the Reference population. A match proportion is the proportion of individuals in the reference population that would result in a match when compared with the crime sample, Y .

A high match proportion would tend to cast doubts on the conviction of the suspect based on the matching criteria while a low match proportion tends to strengthen one's belief in favour of conviction as the observed match has a small chance of being termed as accidental or random. Let us note that the calculation of the match proportion takes place under the assumption of innocence or the null hypotheses H_0 , as one would expect to see in classical statistics. For this reason Berry comments - "A match proportion is an indirect measure of the suspects guilt"

We have noted before that each site contributes 2 readings, one for the mother and one for the father. Let (x_1, x_2) denote the suspect's readings and let (y_1, y_2) denote the values from the crime sample. Then (x_1, y_1) and (x_2, y_2) are compared separately. Let p_i = match proportion for $(x_i, y_i), i = 1, 2$. Then under the assumption of random mating by the Hardy Weinberg law of population genetics the match proportion for the 2 pairs together is taken as $2p_1p_2$. In the homozygous case p is taken as p_1^2 .

Usually matching is done at multiple sites or loci resulting in a match proportion for each site in question. By the assumption of Linkage equilibrium we have independence across sites and so the individual match proportions are multiplied together resulting in the final probability that the Judge gets to see. This P is not a p -value of testing hypothesis but a genuine probability, though calculated under various strong assumptions of independence and applicability of random mating.

Thus we see that conceptually match binning is an easy method - easy to use and understand and explain to others but it has some obvious weaknesses that are discussed as follows : First, the selection of K is arbitrary and depends solely on the subjective judgement of the statistician. Second, there is this arbitrary cut off point - within $Kc\sqrt{2}$ denotes a match but $(K + \epsilon)c\sqrt{2}$ is a non match for any $\epsilon > 0$. Third, there is no way to discriminate between a perfect match and samples that barely match. Both of these cases give rise to the same match proportion - something that seems to be a bit odd. Thus a loss of information takes place in this procedure - the actual observations being

lost to a zero - one scenario - a match or a non-match.

Berry goes on to state the following - “Match Binning is a reasonable inferential method in a scientific setting, and other settings that allow for flexibility but it has several characteristics that make it undesirable for use in court” and goes on to suggest a Bayesian alternative to this classical Match Binning procedure. We discuss his work in the next section. The ideas are not new but can be seen as early as 1977 in D.V. Lindley (1977).

2.3 The Bayesian Formulation as in Berry

We have just seen that in the case of Match Binning probability calculations were done only under H_0 . Since no probability is calculated under the alternative H_1 , to many people, specially Bayesians, the evidence seems incomplete. A Bayesian treats both H_0 and H_1 on the same footing and probability calculations are done under both H_0 and H_1 . He has priors $\pi(\mu, \nu|H_0)$ and $\pi(\mu|H_1)$ under H_0 and H_1 respectively. He calculates the integrated likelihood (also called marginal likelihood) of the data under H_0 and H_1 integrating out the parameters with respect to the corresponding π_i . The ratio of the (integrated) likelihoods, called the Bayes factor, is taken as a measure of evidential strength and is the key factor in decision making in the Bayesian paradigm. The database of fingerprints is used to determine the prior.

Let us define the following :

The Prior Odds of Guilt is the odds of guilt prior to incorporating the DNA evidence. This falls in the judges’ domain and the statistician makes no contribution to this factor.

The Posterior Odds of Guilt is the odds of guilt after the DNA evidence obtained from the crime and suspect samples along with the reference samples have been taken into account. This is the final result that the judge gets to see. The following notations are used throughout : G denotes guilt, I innocence, X denotes the entire DNA evidence i.e. $(x, y, z_1, z_2, \dots, z_n)$, E denotes all other evidence combined. From a simple application of Bayes Theorem we see that the posterior odds of guilt,

$$\begin{aligned} \frac{P(G|X, E)}{P(I|X, E)} &= \frac{P(X|G, E)}{P(X|I, E)} \times \frac{P(G|E)}{P(I|E)} \\ &= \frac{P(X|G)}{P(X|I)} \times \frac{P(G|E)}{P(I|E)} \end{aligned}$$

(as E does not have a contribution in the calculation of the above likelihoods).

= $R \times$ prior odds of Guilt where R is the Bayes factor. The Bayes factor is usually

the summary evidence provided by Bayesians rather than the posterior odds, which is very sensitive to prior odds. Both are the same in objective Bayes Analysis where prior probabilities of both hypotheses are taken to be half.

For a single locus probe the R 's are combined as

$$\begin{aligned} R &= \frac{1}{2}R_{11}R_{22} + \frac{1}{2}R_{12}R_{21} \\ &\sim \frac{1}{2}R_{11}R_{22} \text{ if } x_1 \sim y_1 \\ &\quad \text{and } x_2 \sim y_2 \end{aligned}$$

In case several single locus probes are used the final R is the product of R 's from each single locus probe.

The following distributional assumptions are made :

$$\log x \sim N(\mu, c^2), \log y \sim N(\nu, c^2), \log z_i \sim N(\mu_i, c^2)$$

where (z_1, z_2, \dots, z_n) is the reference dataset, c is a constant. $(x, y, z_1, z_2, \dots, z_n)$ are all independent.

Let

$$\begin{aligned} m &= \frac{\log x + \log y}{2}, s = \frac{|\log x - \log y|}{2} \\ s_i &= \frac{|\log z_i - \log y|}{2} \end{aligned}$$

Berry first calculates the Bayes factor for the Hypothesis of Guilty versus the Hypothesis of Innocence under the following assumptions which are restrictive and unrealistic.

(1) "I" or Innocence implies that someone in the reference population is Guilty, each with prob. $\frac{1}{n}$.

(2) The μ 's are independent with improper uniform prior distribution on $(-\infty, +\infty)$.

$$R = \frac{p(X|\mu = \nu)}{p(X|I)} \text{ as Guilty implies } \mu = \nu.$$

The numerator is

$$\begin{aligned} &\int p(X|\mu = \nu, \mu, \mu_1, \mu_2, \dots, \mu_n) dF(\mu, \mu_1, \dots, \mu_n) \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi c^2 xy} \exp \left\{ -\frac{1}{2c^2} [(\log x - \mu)^2 + (\log y - \mu)^2] \right\} d\mu \\ &\quad \cdot \prod_{i=1}^n \int_{-\infty}^{\infty} \frac{1}{cz_i \sqrt{2\pi}} \exp \left\{ -\frac{1}{2c^2} (\log z_i - \mu_i)^2 \right\} d\mu_i \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi c^2 xy} \int \exp \left\{ -\frac{1}{c^2} [s^2 + (m - \mu)^2] \right\} d\mu \cdot \prod_{i=1}^n \left(\frac{1}{z_i} \right) \\
&= \frac{1}{2cxy\sqrt{\pi}} \exp \left\{ -\left(\frac{s}{c}\right)^2 \right\} \prod_{i=1}^n \frac{1}{z_i} = K \cdot \exp \left\{ -\left(\frac{s}{c}\right)^2 \right\}
\end{aligned}$$

where

$$K = \frac{1}{2cxy\sqrt{\pi}} \cdot \prod_{i=1}^n \frac{1}{z_i}$$

The denominator is $p(X|I)$

$$\begin{aligned}
&= \sum_{i=1}^n p(X | \text{person } i \text{ is guilty}) \cdot P(\text{person } i \text{ guilty} / I) \\
&= \sum_{i=1}^n p(X | \mu_i = \nu) \times \frac{1}{n} \\
&= \frac{K}{n} \sum_{i=1}^n \exp \left\{ -\left(\frac{s_i}{c}\right)^2 \right\}
\end{aligned}$$

Thus the Bayes factor becomes

$$R = \frac{\exp \left\{ -\left(\frac{s}{c}\right)^2 \right\}}{\frac{1}{n} \sum_{i=1}^n \exp \left\{ -\left(\frac{s_i}{c}\right)^2 \right\}}$$

It is a nice and simple formula and helps one to gain insight into the problem, but is based on strong assumptions.

We mention some of the observations that Berry makes based on the above formula. R is large for x and y close together with respect to c . If x and y are far apart then R is small, decreasing exponentially as the square of the log of their ratio. The data points make their contribution through the denominator. If it so happens that x, z_1, z_2, \dots, z_n are all equidistant from y then the ratio is 1 thus exonerating the suspect. This is one scenario where the reference population makes a strong contribution and rightly so. However close x and y may be if all the others exhibit the same closeness to the crime sample one cannot hold the suspect to be guilty. R is close to zero when x and y are far apart compared to one or more samples in the population. The above observations are in agreement with our expectations.

Berry then goes on to work on his second scenario where he assumes the reference data set is a sample from a reference population whose distribution / density may be denoted by $\pi(\mu)$. If $\pi(\mu)$ were known the data points would not have played a role but here they are used to estimate the prior in a frequentist manner.

Here the Bayes factor R is, as before, $\frac{P(X|G)}{P(X|I)}$.

The numerator equals

$$\begin{aligned}
& \int \dots \int p(X|\mu = \nu, \mu, \mu_1, \mu_2, \dots, \mu_n) \pi(\mu, \mu_1, \dots, \mu_n) d\mu d\mu_1 \dots d\mu_n \\
&= \int \dots \int p(X|\mu = \nu, \mu, \mu_1, \mu_2, \dots, \mu_n) \pi(\mu) \pi(\mu_1) \dots \pi(\mu_n) d\mu d\mu_1 \dots d\mu_n. \\
&= \int \dots \int p(x, y|\mu = \nu, \mu) p(z_1|\mu_1) p(z_2|\mu_2) \dots p(z_n|\mu_n) \cdot \pi(\mu) \pi(\mu_1) \\
&\quad \dots \pi(\mu_n) d\mu d\mu_1 \dots d\mu_n \\
&= \int p(x, y|\mu = \nu, \mu) \pi(\mu) d\mu \times \int p(z_1|\mu_1) \pi(\mu_1) d\mu_1 \times \int p(z_2|\mu_2) \pi(\mu_2) d\mu_2 \\
&\quad \times \dots \times \int p(z_n|\mu_n) \pi(\mu_n) d\mu_n
\end{aligned}$$

The denominator equals

$$\begin{aligned}
& \int \dots \int p(X|\mu, \nu, \mu_1, \mu_2, \dots, \mu_n) \pi(\mu, \nu, \mu_1, \dots, \mu_n) d\mu d\nu d\mu_1 \dots d\mu_n \\
&= \int \dots \int p(x, y|\mu, \nu) p(z_1|\mu_1) \dots p(z_n|\mu_n) \pi(\mu) \pi(\nu) \pi(\mu_1) \dots \pi(\mu_n) \\
&\quad d\mu d\nu d\mu_1 \dots d\mu_n \\
&= \int p(x, y|\mu, \nu) \pi(\mu) \pi(\nu) d\mu d\nu \cdot \int p(z_1|\mu_1) \pi(\mu_1) d\mu_1 \dots \int p(z_n|\mu_n) \pi(\mu_n) d\mu_n \\
&= \int p(x|\mu) \pi(\mu) d\mu \cdot \int p(y|\nu) \pi(\nu) d\nu \cdot \int p(z_1|\mu_1) \pi(\mu_1) d\mu_1 \dots \int p(z_n|\mu_n) \pi(\mu_n) d\mu_n.
\end{aligned}$$

Thus we see that the integrals involving $z_1 \dots z_n$ get cancelled from the numerator and denominator and R equals

$$\frac{\int p(x, y|\mu = \nu, \mu) \pi(\mu) d\mu}{\int p(x|\mu) \pi(\mu) d\mu \int p(y|\nu) \pi(\nu) d\nu}$$

The numerator equals

$$\int_{-\infty}^{\infty} \frac{1}{2\pi c^2 xy} \exp \left\{ -\frac{1}{2c^2} [(\log x - \mu)^2 + (\log y - \mu)^2] \right\} \pi(\mu) d\mu$$

and the denominator equals

$$\begin{aligned}
& \int_{-\infty}^{\infty} \frac{1}{c\sqrt{2\pi}x} \exp \left\{ -\frac{1}{2c^2} (\log x - \mu)^2 \right\} \pi(\mu) d\mu \\
& \cdot \int_{-\infty}^{\infty} \frac{1}{c\sqrt{2\pi}y} \exp \left\{ -\frac{1}{2c^2} (\log y - \nu)^2 \right\} \pi(\nu) d\nu
\end{aligned}$$

We are reminded of two important facts here. First, the z_i 's are a sample from the actual population of z 's and hence the sample frequencies do not necessarily reflect the

true population frequencies. And second, z 's are not a sample from the target population represented by $\pi(\mu)$ for the observed values are carrying a measurement error. So the empirical distribution of the $\log z_i$'s cannot be used to estimate the prior. Berry mentions that hierarchical Bayesian methods may be used here but he chooses to adopt a "simple minded density estimation approach". We briefly say what he does and pursue the Empirical Bayes method in the following section.

The second point in the previous paragraph makes the present problem different from standard density estimation. It seems clear that this is a problem of deconvolution and one of the discussants, Chernoff, mentions this. (Berry does not respond to this issue.) To err, if necessary, on the conservative side (i.e. favor the suspect) he chooses an estimate that looks like a kernel estimate and tends to increase the support of the distribution even more than the error model requires. Berry's choice of the density estimate is

$$\widehat{\pi(\mu)} = \frac{1}{n} \sum N(\log z_i, bc^2), \quad b \geq 1.$$

Regarding the choice of b he says "generally, b should be greater than 1 and it should be larger for smaller n . If n is so large that the sample frequencies are effectively the same as the population frequencies, then b could be set to 1."

With the above estimate of the prior Berry calculates the Bayes factor which comes out to

$$R = \frac{Q_2(m) \exp \left\{ - \left(\frac{s}{c} \right)^2 \right\}}{Q_1(\log x) Q_1(\log y)}$$

where

$$Q_j(u) = \frac{1}{n} \sum \frac{1}{\sqrt{1 + jb^2}} \cdot \exp \left\{ - \frac{j}{2c^2} \frac{(u - \log z_i)^2}{1 + jb^2} \right\} \quad j = 1, 2.$$

This expression is far more complicated and the only way to get a feeling for it is to examine the values for different sets of x, y, z_1, \dots, z_n .

Before we end this section we shall see that the match proportion can be looked upon as a Bayes factor. Let the match proportion at a single locus be p . Now p is approximately $P(\text{match} | \text{innocence})$ and $P(\text{match} | \text{guilty})$ is approximately one. So the Bayes factor

$$\frac{P(\text{match} | \text{guilty})}{P(\text{match} | \text{innocence})} = \frac{1}{p} \text{ (approx).}$$

So inference based on p is really inference based on a Bayes factor. Again, this is essentially due to Berry.

2.4 Parametric Empirical Bayes formulation of the DNA fingerprinting problem

We have devoted sections 2.2 and 2.3 to a review of Don Berry's work based on his 1991 paper in Statistical Science. This section onwards we present work that is new. Here we study the DNA fingerprinting problem using the Parametric Empirical Bayes formulation. This was suggested by Prof. Shaw Hwa Lo of Columbia University. Section 2.5 deals with an alternative Bayes factor based on Edwards, Lindman and Savage's (1963) method of enumerating the prior under the null by proper conditioning. Rest of the thesis deals with methodological developments in the area of Empirical Bayes.

Coming back to the topic of this section, we notice that the data is amenable to Empirical Bayes treatment because the database or the reference population consists of observations which are independent and come from a similar environment. Hence by the method of MLII prior estimation the prior parameters and hence the prior can be estimated. We have used the Normal model for the EB development.

The notations used here are as before. The model is as follows :

$$x|\mu \sim N(\mu, c^2)$$

$$y|\nu \sim N(\nu, c^2)$$

$$z_i/\mu_i \sim N(\mu_i, c^2)$$

x, y, z_i 's are all independent and c^2 is assumed known.

$$\mu, \nu, \mu_i \sim N(\eta, \tau^2) \text{ independent}$$

η and τ^2 are not known and are to be estimated from the database or reference population.

We calculate the Bayes factor for the hypothesis of Guilty versus the hypothesis of Innocence.

$$R = \frac{p(X|\mu = \nu)}{p(X|I)} = \frac{p(x, y|\mu = \nu)}{p(x, y|I)}$$

We have shown in Sec. 2.3 that the z terms cancel out from the numerator and denominator and hence do not come into the expression directly.

The Bayes factor is calculated as in Lindley (1977), details of which are shown in Section 2.5. Then the MLII estimates based on the z_i 's are plugged into the expression for the Bayes factor. We note that the estimates for η and τ^2 being based on the z_i 's only, are same for guilt and innocence. The model being a standard one we can use the

estimates as given in Berger (1985), in the final expression.

$$\begin{aligned}
R &= \frac{\int p(x|\mu)p(y|\mu)\pi(\mu)d\mu}{\int p(x|\mu)\pi(\mu)d\mu \cdot \int p(y|\nu)\pi(\nu)d\nu} \\
&= \frac{c^2 + \tau^2}{\sqrt{[2c^2(\tau^2 + c^2/2)]}} \exp \frac{(-d^2\tau^2)}{2c^2(2c^2 + 2\tau^2)} \times \exp \frac{(m - \eta)^2\tau^2}{2(\tau^2 + c^2/2)(c^2 + \tau^2)}
\end{aligned}$$

where

$$M = \frac{X + Y}{2}$$

and

$$D = X - Y$$

We have to replace η and τ^2 by their MLII estimates which are given by

$$\hat{\eta} = \frac{\sum_{i=1}^n z_i}{n}$$

and

$$\hat{\tau}^2 = \max \left(0, \frac{s^2}{n} - c^2 \right)$$

where

$$s^2 = \sum_{i=1}^n (z_i - \hat{\eta})^2$$

Taking into consideration the fact that the prior variance τ^2 will usually be much larger than the error variance c^2 we arrive at an approximated R (as in Lindley 1977) which is

$$\frac{\tau}{\sqrt{2}c} \exp \left(-\frac{d^2}{4c^2} \right) \cdot \exp \left(\frac{(m - \eta)^2}{2\tau^2} \right) \quad (*)$$

The Empirical Bayes Bayes factor is finally obtained by plugging in the EB estimates of η and τ^2 in (*) and it equals

$$\frac{\sqrt{\max \left(0, \frac{s^2}{n} - c^2 \right)}}{\sqrt{2}c} \exp \left(-\frac{d^2}{4c^2} \right) \exp \left(\frac{\left(m - \frac{\sum z_i}{n} \right)^2}{2 \cdot \max \left(0, \frac{s^2}{n} - c^2 \right)} \right)$$

Since $\tau^2 \gg c^2$, we can expect $\frac{s^2}{n} - c^2$ to be positive. Hence the above expression for the

Table 1: Parametric empirical Bayes Bayes factor $\left(\frac{s^2}{n} - c^2\right)^{1/2} / (2^{1/2}c) * \exp(-\frac{1}{2}\lambda^2) * \exp[(m - \bar{z})^2 / \{2(s^2/n - c^2)\}]$ as a function of $\lambda = |x - y| / (2^{1/2}c)$ and $m = \frac{x+y}{2}$ for a sample of size 100, $s^2 = \sum(z_i - \bar{z})^2$, $c = .0042$

$\frac{x+y}{2}$	λ									
	0	0.5	1	1.5	2	2.5	3	4	5	6
2	9061.65	7996.88	5496.17	2941.89	1226.36	398.14	100.67	3.04	0.03	.00014
3	1076.24	949.78	652.77	349.40	145.65	47.29	11.96	0.36	0.004	1.64E-05
3.5	506.89	447.33	307.44	164.56	68.60	22.27	5.63	0.17	0.002	7.72E-06
4	294.0	259.46	178.32	95.45	39.79	12.92	3.27	0.099	0.001	4.48E-06
4.5	210.0	185.33	127.37	68.18	28.42	9.23	2.33	0.07	0.0008	3.2E-06
5	184.73	163.02	112.04	59.97	25.0	8.12	2.05	0.06	0.0007	2.81E-06
5.5	200.11	176.60	121.38	64.97	27.08	8.79	2.22	0.07	0.0007	3.05E-06
6	266.97	235.60	161.92	86.67	36.13	11.73	2.96	0.09	0.001	4.07E-06
6.5	438.60	387.07	266.03	142.39	59.36	19.27	4.87	0.15	0.0016	6.68E-06
7	887.40	783.13	538.24	288.1	120.1	38.99	9.86	0.298	0.003	1.35E-05
8	6784.56	5987.35	4115.04	2202.62	918.19	298.09	75.37	2.28	0.02	.000103

PEB Bayes factor reduces to

$$\frac{\sqrt{\left(\frac{s^2}{n} - c^2\right)}}{\sqrt{2}c} \exp\left(-\frac{d^2}{4c^2}\right) \exp\left(\frac{\left(m - \frac{\sum z_i}{n}\right)^2}{2\left(\frac{s^2}{n} - c^2\right)}\right).$$

A fully Bayesian method of inference would require putting a hierarchical prior on η and τ^2 say $\pi(\eta, \tau^2 / \phi)$ where ϕ is a known hyperparameter (or a vector of known hyperparameters). This method is referred to as Hierarchical Bayes (*HB*). This kind of modelling has become very popular since Markov Chain Monte Carlo (MCMC) was introduced in Bayesian computation.

It is also known, see e.g. Ghosh, Delampady and Samanta (2006, Ch. 9) that in estimation problems both Empirical Bayes and Hierarchical Bayes give very close answers. Usually *EB* is easier to implement but does not take into account the uncertainty due to the estimation of the parameters η and τ^2 . The *HB* method takes care of such problems.

How do the two methods compare for testing problems ? This has received attention only very recently, Scott and Berger (Ann. Statistics, 2010) and elsewhere, e.g. Bogdan, Ghosh and Tokdar (IMS collection vol. 1, 2008). It appears that in testing problems *EB* and *HB* will provide very different answers and *HB* will generally do much better.

2.5 The alternative Bayes factor

In this section we explore the work done by D.V. Lindley (1977) and go on to suggest an alternative Bayes factor. Berry's method, explained in Section 2.3, is in fact based on Lindley's ideas.

Lindley works on the Forensic Identification problem with pieces of glass as his experimental objects and the measurements were taken on the refractive indices of glass pieces. The ideas carry forth to the case of DNA fingerprinting when matching is done at a single locus and for a single pair of observations. He assumes a Normal model for the measurement error and a Normal prior with known parameters. So here the data set z_1, z_2, \dots, z_n or the reference population does not play any role.

We stick to the notations used in previous sections. Thus X denotes the suspect sample and Y the crime sample, the true values of which are μ and ν respectively. X/μ follows the Normal (μ, c^2) distribution and Y/ν follows the Normal (ν, c^2) distribution and are independent. The variance of the measurement error, c^2 , is assumed to be known. μ and ν are assumed to have the normal (η, τ^2) distribution to be denoted by $\pi(\mu)$ and $\pi(\nu)$ respectively and are independent. Then the likelihood Ratio of guilt to innocence can be calculated as follows :

$$R = \frac{p(x, y|G)}{p(x, y|I)}$$

The numerator is calculated as

$$\int p(x|\mu)p(y|\mu)\pi(\mu)d\mu$$

where μ is the common value of the observation under guilt. The denominator is given by

$$\int p(x|\mu)\pi(\mu)d\mu. \int p(y|\nu)\pi(\nu)d\nu$$

as μ and ν can be assumed to be independent under innocence.

The denominator is the product of the unconditional distributions of X and Y and they are both $N(\eta, c^2 + \tau^2)$. The numerator is the joint unconditional distribution of X and Y which is Bivariate Normal with means η , variances $c^2 + \tau^2$ and covariance τ^2 . Transforming to

$$S = X - Y$$

and

$$M = \frac{X + Y}{2},$$

the Jacobian of the transformation being 1, the numerator reduces to the product of the densities of S and M which are $N(0, 2c^2)$ and $N(\eta, \tau^2 + c^2/2)$ respectively.

Similarly the denominator can be written as the product of the densities of S and M which are $N(0, 2c^2 + 2\tau^2)$ and $N(\eta, (c^2 + \tau^2)/2)$ respectively.

The required Likelihood Ratio or the Bayes factor reduces to the following:

$$\frac{c^2 + \tau^2}{\sqrt{[2c^2(\tau^2 + c^2/2)]}} \exp \frac{(-s^2\tau^2)}{2c^2(2c^2 + 2\tau^2)} \exp \frac{(m - \eta)^2\tau^2}{2(\tau^2 + c^2/2)(c^2 + \tau^2)}$$

Thus evidence in Lindley's and Berry's Bayes factor consists of two parts-one coming from comparing density of $X - Y$ under guilt and innocence and another coming from comparing density of $\frac{X+Y}{2}$ under guilt and innocence.

The prior variance τ^2 will usually be much larger than the error variance c^2 . So the above Bayes factor can be approximated by the following expression :

$$\frac{\tau}{\sqrt{2}.c} \exp \left(-\frac{s^2}{4c^2} \right) \cdot \exp \left(\frac{(m - \eta)^2}{2\tau^2} \right)$$

Table 2. Lindley's Bayes factor $\tau(2^{1/2}c)^{-1} \exp \left(-\frac{1}{2}\lambda^2 + \frac{1}{2}\delta^2 \right)$ as a function of $\lambda = |x - y|/(2^{1/2}c)$ and $\delta = \left| \frac{x+y}{2} - \eta \right|/\tau$ for $\tau = 1, c = .0042$ and $\eta = 5$.

λ	δ						
	0	0.5	1.0	1.5	2.0	2.5	3.0
0	168.36	190.77	277.58	518.58	1244.01	3831.83	15155.17
0.5	148.58	168.36	244.96	457.65	1097.84	3381.58	13374.39
1	102.11	115.71	168.36	314.54	754.53	2324.12	9192.08
1.5	54.66	61.94	90.12	168.36	403.87	1244.01	4920.16
2	22.78	25.82	37.57	70.18	168.36	518.58	2051.03
2.5	7.40	8.38	12.20	22.78	54.66	168.36	665.87
3	1.87	2.11	3.08	5.76	13.82	42.57	168.36
4	0.05	0.06	0.09	0.17	0.42	1.28	5.08
5	.00063	.00071	.001	.002	.005	.014	.056
6	2.56E-06	2.91E-06	4.23E-06	7.9E-06	1.89E-05	5.84E-05	.000231

It may be noticed that the first part of the expression is easy to comprehend as it involves the difference of the crime sample and the suspect sample, s .

Larger the difference smaller is the contribution of this term to the Bayes factor and the evidence shifts towards Innocence. Small values of s with respect to c make this

term large thus indicating Guilt. This fits in with our intuitive understanding of the problem as viewed from a classical standpoint. But the significance of the second term is not easy to understand even though we can see that this is the contribution from the prior information and τ being much larger than c it has a much smaller influence on the Bayes factor than does the first term. Some effort at explaining the term would run along the following lines - If x and y are close together then m may be looked upon as the common value of the observations and m close to η would indicate that x and y come from a region with the highest possible density and the contribution to the Bayes factor is small, it being equal to 1 when m exactly equals η . As we move more and more towards unusual values i.e. m is farther away from η the contribution to the Bayes factor increases exponentially. Thus uncommon values of X and Y provide a greater evidence of guilt than does commonly occurring values. But the interpretation is not at all clear when the crime sample and the suspect sample are quite different. In that case m or $(\frac{x+y}{2})$ does not have a proper interpretation nor can the contribution to the Bayes factor be explained. In fact we see that in the case of Innocence the Bayes factor is increased in the direction of guilt - something that we certainly do not want. Our procedure should be such that it gives maximum protection to the innocent but here the M term acts in the opposite direction, however small its impact may be in the case of Innocence. We may notice though that in that case the first term would dominate and the contribution from the second term would be small in comparison.

Thus one may search for an alternative Bayes factor - One that is free of the term involving m , at least in the Normal setup.

We define the prior for μ under guilt, or, equivalently under $\mu = \nu$ by a suitable conditioning on the distribution of μ, ν i.e. as $\pi(\mu|\mu - \nu = 0)$. The calculations for the case of Innocence are the same as in Berry and Lindley's case. Given these priors, the Bayes Factor is calculated as before.

The alternative Bayes factor for a Normal prior :

Model : Same as before

$$\left. \begin{array}{l} \text{Let } U = \mu - \nu \\ \text{And } V = (\mu + \nu)/2 \end{array} \right\} \Rightarrow \begin{array}{l} \mu = V + U/2 \\ \nu = V - U/2 \end{array}$$

U and V are also independent and Normally distributed.

Then

$$\begin{aligned}
& P(a \leq \mu \leq b \mid \mu - \nu) \\
&= P(a \leq V + U/2 \leq b \mid U) \\
&= P(a - U/2 \leq V \leq b - U/2 \mid U) \\
&= P(a - U/2 \leq V \leq b - U/2)
\end{aligned}$$

by the independence of U and V .

This is the unique continuous version of the conditional density of μ given $\mu - \nu$. Setting $U = 0$ in the above leads to $P(a \leq V \leq b)$ which is in fact the distribution of $(\mu + \nu)/2$.

Thus the required prior density in the case of guilt is Normal $(\eta, \tau^2/2)$ instead of the Normal (η, τ^2) density that we were using before. We use the above prior to calculate our alternative Bayes factor. The calculations proceed exactly as before.

$$R = \frac{p(x, y \mid G)}{p(x, y \mid I)}$$

The numerator is calculated as $\int p(x/\mu)p(y/\mu)\pi(\mu/G)d\mu$. The denominator is given by

$$\int p(x/\mu)\pi(\mu)d\mu. \int p(y/\nu)\pi(\nu)d\nu$$

The numerator is the joint unconditional distribution of X and Y which is Bivariate Normal with means η , variances $c^2 + \tau^2/2$ and covariance $\tau^2/2$. The denominator is the product of the unconditional distributions of X and Y , the prior distribution being the $N(\eta, \tau^2)$ distribution as before, and they are both $N(\eta, c^2 + \tau^2)$.

Transforming to $S = X - Y$ and $M = \frac{X + Y}{2}$, the Jacobian of the transformation being 1, the numerator reduces to the product of the densities of S and M which are $N(0, 2c^2)$ and $N(\eta, (c^2 + \tau^2)/2)$ respectively. The denominator is the product of the density of S which is $N(0, 2c^2 + 2\tau^2)$ and the density of M which is $N(\eta, (c^2 + \tau^2)/2)$. Thus the term containing m gets cancelled and we are left with a Bayes factor that involves only s , the difference of the crime sample and the suspect sample. After some simplification we obtain the following expression for the alternative Bayes factor :

$$R = \sqrt{(1 + \tau^2/c^2)} * \exp\left(\frac{-s^2\tau^2}{4c^2(c^2 + \tau^2)}\right)$$

Considering the fact that $\tau \gg c$ we can approximate the above by

$$\sqrt{(1 + \tau^2/c^2)} * \exp(-s^2/4c^2).$$

This formula is much easier to understand and appreciate compared to the existing Bayes factor and we were able to get rid of the confusing presence of the exponential term in M .

We now consider a general non normal prior and proceed in the same way.

Table 3. Alternative Bayes factor $(c^2 + \tau^2)^{1/2}/c * \exp(-\frac{1}{2}\lambda^2)$ as a function of $\lambda = |x - y|/(2^{1/2}c)$ for $\tau = 1$ and $c = .0042$.

λ	Alt. Bayes factor
0	238.0973
0.5	210.1202
1	144.4133
1.5	77.2989
2	32.223
2.5	10.4613
3	2.645
4	0.0799
5	0.0009
6	3.63E-06

The alternative Bayes factor for a general non normal prior :

We follow the same method as before but extend those ideas to non-normal priors. Model : $X|\mu$ follows the Normal (μ, c^2) distribution and $Y|\nu$ follows the Normal (ν, c^2) distribution and are independent. The variance of the measurement error, c^2 , is assumed to be known. Under assumption of independence μ and ν are assumed to have general densities to be denoted by $\pi(\mu)$ and $\pi(\nu)$ respectively.

Let

$$U = \mu - \nu$$

and

$$V = (\mu + \nu)/2$$

as before.

As before we will take $\pi(\mu|\text{Guilt})$ as the conditional distribution of μ given $U = 0$. Let us notice that given Guilt i.e. given $U = 0$ we have $\mu = \nu$ and thus the conditional distribution of $(\mu + \nu)/2$ given Guilt is the same as the conditional distribution of μ given Guilt.

Thus the numerator of the Likelihood Ratio of guilt to innocence which is

$$\int f(x|\mu)f(y|\mu)\pi(\mu|\text{Guilt})d\mu$$

can be expressed as follows :

$$g_1(x-y) \int g_2\left(\frac{x+y}{2}|v\right) \pi(v|U=0)dv$$

where g_1 is the density of $X - Y$ which under guilt is Normal $(0, 2c^2)$, g_2 is the density of $\frac{X+Y}{2}$ given $V = v$ and is Normal $(v, c^2/2)$.

The denominator in the Bayes factor can be written as

$$\int g_1(x-y-u).h_1(u) \left[\int g_2\left(\frac{x+y}{2}-v\right) \pi(v|u)dv \right] du$$

where h_1 is the marginal density of U or $\mu - \nu$ and $\pi^{V|U}(v|u)$ denotes the conditional density of V given U .

In case of independence of U and V , which is true in the case of Normal priors, the Bayes factor reduces to the following :

$$\frac{g_1(x-y)}{\int g_1(x-y-u).h_1(u)du} \quad (**) \quad$$

Even if the above independence of U and V does not hold good we may still use $(**)$ as a possible candidate for the Bayes factor because it is easy to communicate as it depends only on the difference between the crime specimen and the suspect specimen. Second, in the case of unknown prior density, one can compute this without sophisticated density estimation methods that have been discussed in section 2.3. In fact the denominator may be estimated by repeatedly sampling Z_1, Z_2 from the database or the reference population and calculating the empirical distribution of $Z_1 - Z_2$. The required estimate is the density of the smoothed empirical distribution of $Z_1 - Z_2$ at $(x - y)$. The numerator requires no density estimation whatsoever - it is simply the Normal $(0, 2c^2)$ density at $(x - y)$.

Chapter 3

Settling a Conjecture of Morris

3.1 Introduction

This chapter deals with Empirical Bayes Prediction Intervals in a Normal Regression model, using higher order asymptotics. This is joint work with Prof. J.K. Ghosh and Prof. R. Mukerjee.

We consider observable random variables Y_1, Y_2, \dots, Y_k such that given $\theta_1, \theta_2, \dots, \theta_k$, they are independent and Y_j is Normal with mean θ_j and variance $V(1 \leq j \leq k)$. Also $\theta_1, \theta_2, \dots, \theta_k$ are themselves independent and θ_j is normal with mean $z'_j\beta$ and variance $A(1 \leq j \leq k)$. Here $V(> 0)$ is assumed to be known, z'_j is a known $1 \times r$ vector ($1 \leq j \leq k$), β is an $r \times 1$ vector of unknown parameters and $A(> 0)$ is an unknown parameter.

Here we study two popular proposals one due to Morris (1983) and the other discussed in Carlin and Louis (1996, p. 98) for obtaining Empirical Bayes prediction intervals for θ_i on the basis of the observational vector $Y = (Y_1, \dots, Y_k)'$. This is done by developing the higher order asymptotics on the coverage and expected lengths of these intervals, as $k \rightarrow \infty$ keeping i and r fixed. The prediction interval proposed by Morris (1983) is considered in section 3.2, where it is seen that the interval may not attain the target coverage probability in all cases but that a simple modification thereof can achieve this with margin of error $o(k^{-1})$. The intricacies underlying this result are presented in the Appendix. Next, in section 3.3 the idea of Carlin and Louis (1996) is explored analytically to obtain an explicit expression for another prediction interval with the same asymptotic coverage property as the modified version of Morris' interval. In Section 3.4 it is seen that these two intervals, both of which attain the target coverage probability with margin of error $o(k^{-1})$, have the same expected length too up to that order of approximation. It may be emphasized that our implementation of Carlin and Louis (1996) idea in Section 2.3 for the normal regression model is analytic while their discussion seems to be motivated towards numerical derivation of intervals. We refer to Datta *et al.* (2000) for a related result on the choice of noninformative priors in a special case of the setup considered here.

Let Z_k be the $k \times r$ design matrix with rows z'_1, \dots, z'_k . As in Morris (1983), we assume

that $r \leq k - 3$ and that Z_k has full column rank. The following additional assumption is needed.

Assumption. Let $c_{ik} = k\{z'_i(Z'_k Z_k)^{-1} z_i\}$. For fixed i , the sequence $\{c_{ik}\}$ is bounded.

The above assumption holds quite commonly e.g., it holds if the smallest eigenvalue of $k^{-1}Z'_k Z_k$ is bounded away from zero.

3.2 Morris' interval and modification thereof

Let $\xi = (\beta', A)'$ be the vector of unknown parameters and define

$$\left. \begin{aligned} b &= (Z'_k Z_k)^{-1} Z'_k Y, \\ s^2 &= (k - r)^{-1} \sum_{j=1}^k (Y_j - z'_j b)^2, \end{aligned} \right\} \quad (2.1)$$

$$B = V/(V + A), \hat{B} = \frac{k - r - 2}{k - r} \cdot \frac{V}{\max(V, s^2)}, \quad (2.2)$$

$$\rho = B/(1 - B), \hat{\rho} = \hat{B}/(1 - \hat{B}), \quad (2.3)$$

$$h_{ik}(u; \rho) = \frac{1}{2}(c_{ik} - r)\rho + \frac{1}{4}(u^2 + 1)\rho^2 \quad (-\infty < u < \infty) \quad (2.4)$$

$$\hat{\theta}_i = (1 - \hat{B})Y_i + \hat{B}z'_i b, \quad (2.5)$$

$$s_i^2 = V(1 - \frac{k - r}{k}\hat{B}) + \frac{2}{k - r - 2}\hat{B}^2(Y_i - z'_i b)^2. \quad (2.6)$$

With a target coverage probability of at least $1 - \alpha$, Morris (1983) proposed the prediction interval $\hat{\theta}_i \pm z s_i$ for θ_i , where z is the upper $\alpha/2$ point of a standard normal variate. We quote from Morris (1983, pp. 51-52). “Does (4.3) have confidence coefficient at least $1 - \alpha$ for every $K \geq r + 3$ and all $\beta \in R^r$, $A \geq 0$? There is mathematical and numerical evidence (Morris 1981, and unpublished) that this is so for the usual levels of $1 - \alpha$. No formal proof has yet been given, but if the coverage probability of (4.3) ever is less than $1 - \alpha$, it could be only by a negligible amount.”

The following theorem, proved in the appendix, is crucial in studying this interval.

Theorem 2.1. For any convergent non stochastic sequence $\{t_k\}$ of real numbers,

$$P_\xi\{(\theta_i - \hat{\theta}_i)/s_i \leq t_k\} = \Phi(t_k) - k^{-1}t_k\phi(t_k)h_{ik}(t_k; \rho) + o(k^{-1}),$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, respectively.

For Morris' (1983) interval $\hat{\theta}_i \pm z s_i$, by (2.4) and Theorem 2.1,

$$P_\xi(\hat{\theta}_i - z s_i \leq \theta_i \leq \hat{\theta}_i + z s_i) = 1 - \alpha - 2k^{-1}z\phi(z)h_{ik}(z; \rho) + o(k^{-1}). \quad (2.7)$$

Interestingly, $h_{ik}(z; \rho)$ can be positive and hence the term of order $O(k^{-1})$ in (2.7) can be negative. This happens, for example if $r = 1$ and Z_k equals the $k \times 1$ vector of 1's (cf Datta *et al.* 2000). Then $c_{ik} = 1$ and by (2.4), $h_{ik}(z; \rho) = \frac{1}{4}(z^2 + 1)\rho^2$. Thus the coverage probability of the interval $\hat{\theta}_i \pm z s_i$, as it stands, can fall short of $1 - \alpha$ upto the order of approximation considered in (2.7), i.e., Morris' interval is not only not probability matching (as pointed out in the next chapter) but is also not conservative upto $o(\frac{1}{k})$.

We now indicate a simple modification of Morris' (1983) interval that attains a coverage probability $1 - \alpha$ with margin of error $o(k^{-1})$. This is given by

$$I = \left[\hat{\theta}_i - z \{1 + k^{-1}h_{ik}(z; \hat{\rho})\} s_i, \hat{\theta}_i + z \{1 + k^{-1}h_{ik}(z; \hat{\rho})\} s_i \right]. \quad (2.8)$$

By Eq. (A.7) in the appendix, $\hat{B} = B + o(1)$ and hence $\hat{\rho} = \rho + o(1)$, on a set with P_ξ -probability $1 + o(k^{-1})$ uniformly over compact ξ -sets. Hence by (2.8) and Theorem 2.1,

$$\begin{aligned} P_\xi(\theta_i \in I) &= P_\xi \left[-z \{1 + k^{-1}h_{ik}(z; \rho)\} \right. \\ &\quad \left. \leq (\theta_i - \hat{\theta}_i)/s_i \leq z \{1 + k^{-1}h_{ik}(z; \rho)\} \right] + o(k^{-1}) \\ &= 2\Phi(z) - 1 + o(k^{-1}) = 1 - \alpha + o(k^{-1}). \end{aligned}$$

In fact, by Theorem 2.1, it can similarly be seen that

$$P_\xi \left[\theta_i > \hat{\theta}_i + z \{1 + k^{-1}h_{ik}(z; \hat{\rho})\} s_i \right] = \alpha/2 + o(k^{-1}),$$

$$P_\xi \left[\theta_i < \hat{\theta}_i - z \{1 + k^{-1}h_{ik}(z; \hat{\rho})\} s_i \right] = \alpha/2 + o(k^{-1}).$$

Hence, in addition to attaining a coverage probability $1 - \alpha$ with margin of error $o(k^{-1})$, the interval I is equal tailed up to the same order of approximation.

3.3 Another interval

We now explore analytically the idea of Carlin and Louis (1996) to get another prediction interval for θ_i . Observe that given Y conditionally θ_i is normal $N((1 - B)Y_i +$

$Bz'_i\beta, V(1 - B))$. Hence defining $\theta_i^* = (1 - B^*)Y_i + B^*z'_ib$, where

$$B^* = V / \max \left(V, \frac{k - r}{k} s^2 \right)$$

is the maximum likelihood estimator (MLE) of B , one can check that this approach yields a prediction interval for θ_i , of the form

$$\left[\theta_i^* + t^{(1)}(\xi^*) \{V(1 - B^*)\}^{1/2}, \theta_i^* + t^{(2)}(\xi^*) \{V(1 - B^*)\}^{1/2} \right].$$

In the above, ξ^* is the MLE of ξ and $t^{(1)}(\cdot)$ and $t^{(2)}(\cdot)$ are such that

$$P_\xi \left[(\theta_i - \theta_i^*) / \{V(1 - B^*)\}^{1/2} < t^{(1)}(\xi) \right] = \alpha/2, \quad (3.1)$$

$$P_\xi \left[(\theta_i - \theta_i^*) / \{V(1 - B^*)\}^{1/2} > t^{(2)}(\xi) \right] = \alpha/2. \quad (3.2)$$

The following theorem helps in the approximate determination of $t^{(1)}(\cdot)$ and $t^{(2)}(\cdot)$ so that (3.1) and (3.2) hold with margin of error $o(k^{-1})$. Its proof is similar to that of Theorem 2.1 and hence omitted. In what follows, for any real u ,

$$h_{ik}^*(u; \rho) = \frac{1}{2}(c_{ik} + r + 4)\rho + \frac{1}{4}(u^2 + 1)\rho^2. \quad (3.3)$$

Theorem 3.1. *For any convergent nonstochastic sequence $\{t_k\}$ of real numbers,*

$$P_\xi \{ (\theta_i - \theta_i^*) / \{V(1 - B^*)\}^{1/2} \leq t_k \} = \Phi(t_k) - k^{-1}t_k\phi(t_k)h_{ik}^*(t_k; \rho) + o(k^{-1}).$$

By Theorem 3.1, with

$$t^{(1)}(\xi) = -z\{1 + k^{-1}h_{ik}^*(z; \rho)\}, \quad t^{(2)}(\xi) = z\{1 + k^{-1}h_{ik}^*(z; \rho)\},$$

conditions (3.1) and (3.2) hold with margin of error $o(k^{-1})$. Hence the idea of Carlin and Louis (1996) yields the prediction interval

$$I^* = \left[\theta_i^* - z\{1 + k^{-1}h_{ik}^*(z; \rho^*)\}\{V(1 - B^*)\}^{1/2}, \theta_i^* + z\{1 + k^{-1}h_{ik}^*(z; \rho^*)\}\{V(1 - B^*)\}^{1/2} \right]$$

for θ_i , where ρ^* equals $B^*/(1 - B^*)$ if $B^* < 1$, and is defined arbitrarily if $B^* = 1$. As in Section 3.2, by Theorem 3.1, this interval is equal tailed up to $o(k^{-1})$ in addition to attaining a coverage probability $1 - \alpha$ up to the same order of approximation.

3.4 Expected lengths

Let L and L^* be the lengths of the intervals I and I^* , respectively. Define

$$Q_{ik}(z; \rho) = \left(\frac{1}{2}c_{ik} + 1 \right) \rho + \frac{1}{4}z^2\rho^2.$$

Then the following theorem holds.

Theorem 4.1. (a) $E_\xi(L) = 2z\{V(1-B)\}^{1/2}\{1 + k^{-1}Q_{ik}(z; \rho)\} + o(k^{-1})$.

(b) $E_\xi(L^*) = 2z\{V(1-B)\}^{1/2}\{1 + k^{-1}Q_{ik}(z; \rho)\} + o(k^{-1})$.

Thus the two prediction intervals I and I^* , based respectively on modification of Morris' (1983) proposal and analytic implementation of Carlin and Louis' (1996) idea, are at par with respect to both coverage probability and expected length, even under higher order asymptotics retaining terms of order $O(k^{-1})$.

Chapter 4

Rigorous Justification of the Prior of Datta *et al.* (2000)

4.1 Introduction

In the previous chapter we have examined a conjectured inequality in Morris (1983). Here we will study the choice of prior in the manner of Datta *et al.* (2000). Our main contribution is to point out an error in their derivation and indicate how this can be rectified. Actually it amounts to a different proof of their interesting result.

In the next section we provide a background and overview as well as summarize the main facts in Datta *et al.* (2000). We point out why it has what seems to be a nonrectifiable error and in Section 4.3 rederive their result. This is joint work with Prof. J.K. Ghosh and Prof. Rahul Mukerjee.

4.2 Review and Main result of Datta *et al.* (2000).

We use a simplified version of the Model used in Chapter 3. The notations have to be changed slightly for computational convenience.

Given $\theta_1, \theta_2, \dots, \theta_n$, the random variables Y_1, Y_2, \dots, Y_n are independent and Y_i follows Normal (θ_i, V) , $i = 1, 2, \dots, n$. Also $\theta_1, \dots, \theta_n$ are iid such that for each i , θ_i is Normal with mean ξ_1 and variance ξ_2 ($-\infty < \xi_1 < \infty$, $\xi_2 > 0$). Here V is known while ξ_1, ξ_2 are unknown parameters. Assuming V to be known is not unrealistic as a good estimate of V based on many degrees of freedom is usually available.

The PEB estimates of ξ_1 , $B = \frac{V}{V+\xi_2}$ and θ_i are

$$\begin{aligned}\hat{\xi}_1 &= \bar{y} \\ \hat{B} &= (n-3)\frac{V}{S} \text{ with } S = \sum (y_i - \bar{y})^2 \\ \hat{\theta}_i &= (1 - \hat{B})y_i + \hat{B}\bar{y}\end{aligned}$$

and estimate of posterior variance of θ_i , namely, $V(1 - B)$ is $V(1 - \hat{B})$.

The main interest is in PEB confidence intervals for θ_i , of size $1 - \alpha$, i.e. random sets

CI depending on y_1, y_2, \dots, y_n only and satisfying the conservative PEB requirement

$$P_{\xi_1, \xi_2}(\theta_i \in CI) \geq 1 - \alpha \quad \forall \xi_1, \xi_2 \quad (2.1)$$

It is known from standard asymptotics that for $CI = \hat{\theta}_i \pm z_{\alpha/2} \sqrt{V(1 - \hat{B})}$, the naive confidence interval,

$$P_{\xi_1, \xi_2}(\theta_i \in CI) = 1 - \alpha + O(n^{-1}) \quad \forall \xi_1, \xi_2 \quad (2.2)$$

This is true for virtually any prior satisfying certain regularity conditions. For reasons explained later the naive interval is not expected to satisfy (2.1) but does satisfy (2.2). One of the goals in the previous chapter and here is to strengthen (2.2) to

$$P_{\xi_1, \xi_2}(\theta_i \in CI) = 1 - \alpha + o(n^{-1}) \quad (2.3)$$

The $o(n^{-1})$ term is actually $O(n^{-2})$. Though we have not checked, we believe it is not hard to show that (2.3) fails for the naive PEB interval. We showed in the previous chapter (2.3) fails for Morris's interval also.

Morris points out, what is now well known, but not so in 1983, that the naive PEB variance estimate, $V(1 - \hat{B})$, is too small because it does not take into account the variability of ξ_1 and $B = \frac{V}{V + \xi_2}$, which are treated as constants even though they are actually estimated by \bar{y}, \hat{B} .

Morris (1981) then uses Bayesian theory “to obtain a more complete analysis”. He puts priors on ξ_1, ξ_2 - both uniform - and wished to study the frequentist properties of his credibility intervals. Morris proposes a confidence interval $\hat{\theta}_i \pm z_{\alpha/2} \cdot s_i$ where, as in Chapter 3, with $r = 1$,

$$s_i^2 = V \left(1 - \frac{n-1}{n} \hat{B} \right) + \frac{2}{n-3} \hat{B}^2 (y_i - \bar{y})^2$$

Morris probably does this with the expectation that some probability matching will take place and (2.1) may hold, even though he never mentions anything about probability matching.

There has been a lot of recent work on the choice of priors which has some frequentist validation in the sense that a posterior quantile of the parameter of interest should equal the corresponding frequentist probability up to a suitable order. This condition leads to a differential equation for the prior. Such work dates back to Welch and Peers (1963),

Stein (1985), Tibshirani (1989), J.K. Ghosh (1994), Diccio and Stern (1993, 94), J.K. Ghosh and Mukerjee (1991, 92, 93a, b, 94a, b, 95a, b), Datta and M. Ghosh (1995a, b) etc. A recent Monograph with a review and many new results in this area is Datta and Mukerjee (2004).

Datta, M. Ghosh and Mukerjee (2000) showed that Morris's priors are not probability matching. They find a correct probability matching prior - but the proof is incorrect. We provide a corrected proof in the next section.

We end this section with a brief discussion of why the proof of Datta *et al.* has an error.

Datta *et al.* confine attention to a prior of the form

$$\pi(\xi_1, \xi_2) = \text{const.} \pi(\xi_2)$$

i.e. ξ_1 and ξ_2 are independent and the prior for ξ_1 is taken to be the Lebesgue measure. Morris's prior is of this form.

Since the prior for ξ_1 is already chosen, ξ_1 is integrated and the resulting density is denoted by P_{ξ_2} . Then the probability matching equation is written as

$$P_{\xi_2} \left\{ \theta_i \in \left(\hat{\theta}_{\alpha/2, i}, \hat{\theta}_{1-\alpha/2, i} \right) \right\} = 1 - \alpha + o(n^{-1}) \quad \forall \xi_2 \quad (2.4)$$

where $\hat{\theta}_{\alpha/2, i}, \hat{\theta}_{1-\alpha/2, i}$ are the posterior quantiles i.e.

$$\pi \left\{ \theta_i \in \left(\hat{\theta}_{\alpha/2, i}, \hat{\theta}_{1-\alpha/2, i} \right) | y \right\} = 1 - \alpha \quad (2.5)$$

They then get a differential equation for $\pi(\xi_2)$ using a "Bayesian route to Frequentist calculations" discussed in Datta and Mukerjee (2004). Unfortunately neither (2.4) nor the Bayesian Route can be justified because the prior for ξ_1 is improper. This implies the L.H.S. of (2.4) is infinity.

In the next section a correct proof is presented briefly.

4.3 Modified derivation of the result in Section 3 of Datta, Ghosh and Mukerjee (2000)

The normal hierarchical Bayes model has been stated before but we mention it here again.

$$Y_i | \theta_i, V \stackrel{\text{indep}}{\sim} N(\theta_i, V), \quad i = 1, 2, \dots, n$$

$$\theta_i | \xi_1, \xi_2 \stackrel{\text{iid}}{\sim} N(\xi_1, \xi_2), \quad i = 1, 2, \dots, n \quad -\infty < \xi_1 < \infty, \xi_2 > 0.$$

V is assumed known.

We observe that marginally (i.e. integrating $\theta_1, \theta_2, \dots, \theta_n$ out) Y_1, \dots, Y_n are iid, each normal with mean ξ_1 and variance $V + \xi_2$. Thus the joint density of Y_1, \dots, Y_n is $e^{nl(\xi)}$ where $\xi = (\xi_1, \xi_2)'$ and

$$l(\xi) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log(V + \xi_2) - \frac{1}{2n(V + \xi_2)} \sum_{j=1}^n (y_j - \xi_1)^2$$

Let $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2)$ be the MLE of ξ based on $l(\xi)$. Define

$$D_j = \partial / \partial \xi_j, \quad a_{jr} = D_j D_r l(\xi) |_{\hat{\xi}}, \quad c_{jr} = -a_{jr}$$

$$a_{jrs} = D_j D_r D_s l(\xi) |_{\hat{\xi}}, \quad C = ((c_{jr})), \quad C^{-1} = ((c^{jr}))$$

The formal expansions considered below are valid over a set S having P_ξ -probability $1 + o(n^{-1})$ uniformly over compact sets of ξ . The matrix C is positive definite over S .

The joint density of θ_i and Y_1, \dots, Y_n is $(n+1)$ -variate normal with mean vector $\xi_1 1_{n+1}$ and dispersion matrix

$$\begin{pmatrix} \xi_2 & \xi_2 e_i' \\ \xi_2 e_i & (V + \xi_2) I_n \end{pmatrix}$$

where 1_{n+1} is the $(n+1) \times 1$ vector with all elements unity, I_n is the $n \times n$ identity matrix and e_i is the $n \times 1$ vector with 1 at i^{th} position and 0 elsewhere. Hence for fixed ξ , the conditional distribution of θ_i given Y_1, \dots, Y_n is normal with mean

$$\mu(\xi) = \xi_1 + \frac{\xi_2}{V + \xi_2} (Y_i - \xi_1) = \frac{V \xi_1}{V + \xi_2} + \frac{\xi_2 Y_i}{V + \xi_2}$$

and variance $\lambda(\xi) = \frac{V \xi_2}{V + \xi_2}$.

Therefore, the joint density of θ_i and Y_1, \dots, Y_n is $f(\theta_i, \mu(\xi), \lambda(\xi)) \exp\{nl(\xi)\}$ where $f(., \mu(\xi), \lambda(\xi))$ is the univariate normal density with mean $\mu(\xi)$ and variance $\lambda(\xi)$.

Posterior predictive density of θ_i

Let us consider a prior density $\pi(\xi)$ which is positive and thrice continuously differentiable. The posterior predictive density of θ_i , given $Y = (Y_1, \dots, Y_n)'$, under the prior

$\pi(\xi)$ is

$$\begin{aligned}\pi(\theta_i|Y) &= \frac{\int b(\xi)e^{nl(\xi)}\pi(\xi)d\xi}{\int e^{nl(\xi)}\pi(\xi)d\xi} \\ &= \frac{\int b(\xi)e^{n\{l(\xi)-l(\hat{\xi})\}}\pi(\xi)d\xi}{\int e^{n\{l(\xi)-l(\hat{\xi})\}}\pi(\xi)d\xi}\end{aligned}$$

where

$$b(\xi) = f(\theta_i, \mu(\xi), \lambda(\xi)).$$

We write

$$h = (h_1, h_2)' = \sqrt{n}(\xi - \hat{\xi}).$$

Then the above yields

$$\pi(\theta_i|Y) = \frac{\int b\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) \exp\left[n\left\{l\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) - l(\hat{\xi})\right\}\right] \pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) dh}{\int \exp\left[n\left\{l\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) - l(\hat{\xi})\right\}\right] \pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) dh}$$

We write

$$\begin{aligned}\hat{b} &= b(\hat{\xi}), \hat{b}_j = D_j b(\xi)|_{\hat{\xi}}, \hat{b}_{jr} = D_j D_r b(\xi)|_{\hat{\xi}}, \\ \hat{\pi} &= \pi(\hat{\xi}), \hat{\pi}_j = D_j \pi(\xi)|_{\hat{\xi}}, \hat{\pi}_{jr} = D_j D_r \pi(\xi)|_{\hat{\xi}},\end{aligned}$$

Using summation convention,

$$b\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) = \hat{b} + n^{-1/2}h_j\hat{b}_j + \frac{1}{2}n^{-1}h_jh_r\hat{b}_{jr} + o(n^{-1})$$

Since

$$\begin{aligned}&\int e^{n\{l(\hat{\xi} + \frac{h}{\sqrt{n}}) - l(\hat{\xi})\}} \pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) dh \\ &= \frac{\hat{\pi}(2Pi)}{|C|^{1/2}} + o(n^{-1/2})\end{aligned}$$

[Vide Ghosh and Mukerjee (1991); here Pi means the number π .] it follows that

$$\begin{aligned}\pi(\theta_i|Y) &= \frac{\int \left[\hat{b} + n^{-1/2}h_j\hat{b}_j + \frac{n^{-1}}{2}h_jh_r\hat{b}_{jr}\right] e^{n\{l(\hat{\xi} + \frac{h}{\sqrt{n}}) - l(\hat{\xi})\}} \pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) dh}{\int e^{n\{l(\hat{\xi} + \frac{h}{\sqrt{n}}) - l(\hat{\xi})\}} \pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) dh} + o(n^{-1}) \\ &= \hat{b} + \frac{\int \left[n^{-1/2}h_j\hat{b}_j + \frac{1}{2}n^{-1}h_jh_r\hat{b}_{jr}\right] e^{n\{l(\hat{\xi} + \frac{h}{\sqrt{n}}) - l(\hat{\xi})\}} \pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) dh}{\hat{\pi}(2Pi)/|C|^{1/2}} + o(n^{-1})\end{aligned}$$

Now

$$\begin{aligned}
\pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) &= \hat{\pi} + \frac{1}{\sqrt{n}}h_j\hat{\pi}_j + o(n^{-1/2}), \\
n\left\{l\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) - l(\hat{\xi})\right\} &= -\frac{1}{2}h'Ch + \frac{1}{6\sqrt{n}}h_jh_rh_sa_{jrs} + o(n^{-1/2}), \\
\exp\left[n\left\{l\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) - l(\hat{\xi})\right\}\right] &= e^{-\frac{1}{2}h'Ch}\left[1 + \frac{1}{6\sqrt{n}}h_jh_rh_sa_{jrs}\right] + o(n^{-1/2}), \\
\left[n^{-1/2}h_j\hat{b}_j + \frac{1}{2}n^{-1}h_jh_r\hat{b}_{jr}\right] &\left[\exp\left(n\left\{l\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) - l(\hat{\xi})\right\}\right)\right]\pi\left(\hat{\xi} + \frac{h}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{n}}\hat{\pi}e^{-\frac{1}{2}h'Ch}\left[h_j\hat{b}_j + \frac{1}{2\sqrt{n}}h_jh_r\hat{b}_{jr}\right]\left[1 + \frac{1}{6\sqrt{n}}h_jh_rh_sa_{jrs}\right]\left[1 + \frac{1}{\sqrt{n}}h_j\frac{\hat{\pi}_j}{\hat{\pi}}\right] \\
&\quad + o(n^{-1}) \\
&= \frac{1}{\sqrt{n}}\hat{\pi}e^{-\frac{1}{2}h'Ch}\left[h_j\hat{b}_j + \frac{1}{\sqrt{n}}\left(\frac{1}{2}h_jh_r\hat{b}_{jr} + \frac{1}{6}h_jh_rh_s h_t a_{jrs}\hat{b}_t + h_jh_r\frac{\hat{\pi}_j}{\hat{\pi}}\hat{b}_r\right)\right] + o(n^{-1})
\end{aligned}$$

Hence writing $\phi_2(\cdot; C^{-1})$ for the bivariate normal density with null mean vector and dispersion matrix C^{-1} .

$$\begin{aligned}
\pi(\theta_i|Y) &= \hat{b} + \frac{1}{\sqrt{n}}\int \phi_2(h; C^{-1})\left[h_j\hat{b}_j + \frac{1}{\sqrt{n}}\left(\frac{1}{2}h_jh_r\hat{b}_{jr} + \frac{1}{6}h_jh_rh_s h_t a_{jrs}\hat{b}_t\right.\right. \\
&\quad \left.\left.+ \frac{h_jh_r\hat{\pi}_j\hat{b}_r}{\hat{\pi}}\right)\right]dh + o(n^{-1})
\end{aligned} \tag{3.1}$$

The last step follows as a_{jrs} is invariant under permutation of its subscripts.

Frequentist validity of the posterior quantiles of θ_i

By (3.1), the $(1 - \alpha)^{th}$ posterior quantile of θ_i is

$$Q(\pi) = q(\hat{\xi}) + \frac{u}{n}$$

where

$$q(\hat{\xi}) = \mu(\hat{\xi}) + z\sqrt{\lambda(\hat{\xi})}$$

z is the $(1 - \alpha)^{th}$ quantile of a standard normal variate, and u may depend on the prior π , in addition to α and Y , and is at most of order $O(1)$.

We now consider an auxiliary prior $\bar{\pi}(\xi)$ and from (3.1) we note that

$$\bar{\pi}(\theta_i|Y) = \pi(\theta_i|Y) + \frac{1}{n}\left(\frac{\hat{\pi}_j}{\hat{\pi}} - \frac{\hat{\pi}_j}{\hat{\pi}}\right)c^{jr}\hat{b}_r + o(n^{-1})$$

Hence writing $P^{\bar{\pi}}\{.\mid Y\}$ for the posterior probability measure under the prior $\bar{\pi}(\cdot)$

$$P^{\bar{\pi}}\{\theta_i \leq Q(\pi) \mid Y\} = 1 - \alpha + \frac{1}{n} \left(\frac{\widehat{\pi}_j}{\widehat{\pi}} - \frac{\widehat{\pi}_j}{\widehat{\pi}} \right) c^{jr} \int_{-\infty}^{q(\widehat{\xi})} \widehat{b}_r d\theta_i + o(n^{-1})$$

We next note that

$$\begin{aligned} \log b(\xi) &= \log f(\theta_i, \mu(\xi), \lambda(\xi)) \\ &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \lambda(\xi) - \frac{1}{2\lambda(\xi)} \{\theta_i - \mu(\xi)\}^2 \\ \frac{D_1 b(\xi)}{b(\xi)} &= \frac{1}{2\lambda(\xi)} 2\{\theta_i - \mu(\xi)\} \{D_1 \mu(\xi)\} = \frac{\theta_i - \mu(\xi)}{\lambda(\xi)} \left(\frac{V}{V + \xi_2} \right) \\ \frac{D_2 b(\xi)}{b(\xi)} &= -\frac{D_2 \lambda(\xi)}{2\lambda(\xi)} + \frac{D_2 \lambda(\xi)}{2\{\lambda(\xi)\}^2} \{\theta_i - \mu(\xi)\}^2 \\ &\quad + \frac{1}{2\lambda(\xi)} 2\{\theta_i - \mu(\xi)\} \{D_2 \mu(\xi)\} \\ &= \left(\frac{D_2 \lambda(\xi)}{2\lambda(\xi)} \right) \left[\frac{\{\theta_i - \mu(\xi)\}^2}{\lambda(\xi)} - 1 \right] + \frac{\theta_i - \mu(\xi)}{\lambda(\xi)} \frac{V}{(V + \xi_2)^2} (Y_i - \xi_1) \\ &= \frac{V}{2\xi_2(V + \xi_2)} \left[\frac{\{\theta_i - \mu(\xi)\}^2}{\lambda(\xi)} - 1 \right] + \frac{\theta_i - \mu(\xi)}{\lambda(\xi)} \cdot \frac{V(Y_i - \xi_1)}{(V + \xi_2)^2} \end{aligned}$$

Therefore, writing $\phi(\cdot)$ for the standard univariate normal density, we have

$$\begin{aligned} \int_{-\infty}^{q(\widehat{\xi})} \widehat{b}_1 d\theta_i &= \int_{-\infty}^z \frac{w}{\sqrt{\lambda(\widehat{\xi})}} \left(\frac{V}{V + \widehat{\xi}_2} \right) \phi(w) dw \\ &= -\sqrt{\frac{V}{\widehat{\xi}_2(V + \widehat{\xi}_2)}} \phi(z), \\ \int_{-\infty}^{q(\widehat{\xi})} \widehat{b}_2 d\theta_i &= \int_{-\infty}^z \left[\frac{V}{2\widehat{\xi}_2(V + \widehat{\xi}_2)} (w^2 - 1) + \frac{V(Y_i - \widehat{\xi}_1)}{(V + \widehat{\xi}_2)^2 \sqrt{\lambda(\widehat{\xi})}} w \right] \phi(w) dw \\ &= -\frac{V}{2\widehat{\xi}_2(V + \widehat{\xi}_2)} z \phi(z) - \frac{V(Y_i - \widehat{\xi}_1)}{(V + \widehat{\xi}_2)^2 \sqrt{\lambda(\widehat{\xi})}} \phi(z). \end{aligned}$$

Also, it is easy to see that

$$c^{11} = V + \widehat{\xi}_2, \quad c^{12} = c^{21} = 0,$$

$$c^{22} = \left[\frac{1}{n(V + \widehat{\xi}_2)^3} \sum_{j=1}^n (Y_j - \widehat{\xi}_1)^2 - \frac{1}{2(V + \widehat{\xi}_2)^2} \right]^{-1}$$

Since $\hat{\xi}_1 = \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j$ and

$$\hat{\xi}_2 = \max \left[\frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2 - V, 0 \right],$$

from the above it follows that

$$\begin{aligned} & E_{\xi} \left[P^{\bar{\pi}} \{ \theta_i \leq Q(\pi) | Y \} \right] \\ = & 1 - \alpha - \frac{1}{n} \left[\left\{ \frac{\bar{\pi}_1(\xi)}{\bar{\pi}(\xi)} - \frac{\pi_1(\xi)}{\pi(\xi)} \right\} (V + \xi_2) A_1(\xi) \right. \\ & \left. + \left\{ \frac{\bar{\pi}_2(\xi)}{\bar{\pi}(\xi)} - \frac{\pi_2(\xi)}{\pi(\xi)} \right\} \{ 2(V + \xi_2)^2 \} A_2(\xi) \right] + o(n^{-1}) \end{aligned}$$

where

$$\begin{aligned} A_1(\xi) &= \sqrt{\frac{V}{\xi_2(V + \xi_2)}} \phi(z), \\ A_2(\xi) &= \frac{V}{2\xi_2(V + \xi_2)} z \phi(z). \end{aligned}$$

Hence, using the shrinkage argument,

$$\begin{aligned} P_{\xi}(\theta_i \leq Q(\pi)) &= 1 - \alpha + \frac{1}{n} \left[\frac{\pi_1(\xi)}{\pi(\xi)} (V + \xi_2) A_1(\xi) + D_1 \{ (V + \xi_2) A_1(\xi) \} \right. \\ &\quad \left. + \frac{\pi_2(\xi)}{\pi(\xi)} \{ 2(V + \xi_2)^2 \} A_2(\xi) + D_2 \{ 2(V + \xi_2)^2 A_2(\xi) \} \right] + o(n^{-1}) \\ &= 1 - \alpha + \frac{1}{n\pi(\xi)} \left[D_1 \{ \pi(\xi)(V + \xi_2) A_1(\xi) \} \right. \\ &\quad \left. + D_2 \{ \pi(\xi)(2(V + \xi_2)^2) A_2(\xi) \} \right] + o(n^{-1}) \end{aligned}$$

Hence, the matching condition becomes

$$D_1 [\pi(\xi)(V + \xi_2) A_1(\xi)] + D_2 [\pi(\xi) \{ 2(V + \xi_2)^2 \} A_2(\xi)] = 0$$

suppose $\pi(\xi) = \pi(\xi_2)$ i.e. $\pi(\xi)$ depends on ξ only through ξ_2 . Then $\pi(\xi)$ satisfies the above matching condition if and only if

$$D_2 [\pi(\xi) \{ 2(V + \xi_2)^2 \} A_2(\xi)] = 0$$

i.e. iff

$$D_2 \left[\pi(\xi)(V + \xi_2)^2 \frac{1}{\xi_2(V + \xi_2)} \right] = 0$$

i.e. iff

$$\pi(\xi) \propto \frac{\xi_2}{V + \xi_2},$$

which is the same as the prior reported in page 188 of Datta *et al.* (2000).

References

- [1] BASU, R., GHOSH, J.K. AND MUKHERJEE, R. (2003). Empirical Bayes prediction intervals in a normal regression model : higher order asymptotics. *Statistics & Probability Letters*, **63**, 197-203.
- [2] BERGER, J.O. (1986). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.
- [3] BERRY, D.A. (1991). Inferences using DNA Profiling in Forensic Identification and Paternity Cases. *Statistical Science*, **6**, 2 : 175-205.
- [4] BERRY, D.A., EVETT, I.W. AND PINCHIN, R. (1991). Statistical inference in crime investigations using DNA profiling. *J. Roy. Statist. Soc., Ser. C*.
- [5] BERRY, D.A., EVETT, I.W. AND PINCHIN, R. (1992). Statistical Inference in Crime Investigations using Deoxyribonucleic Acid Profiling. *Appl. Statist.*, **41**, No. 3, 499-531.
- [6] BICKEL, PETER J. AND GHOSH, J.K. (1990). A Decomposition for the likelihood Ratio Statistic and the Bartlett Correction - A Bayesian Argument. *Annals. of Statistics*, Vol. **18**, No. 3.
- [7] BODMER, WALTER AND MCKIE, ROBIN (1994). *The Book of Man : The Quest to Discover our Genetic Heritage*. Little Brown and Co., London.
- [8] BOGDAN, M., GHOSH, J.K. AND TOKDAR, S.T. (2008). A comparison of the Benjamini - Hochberg procedure with some Bayesian rules for multiple testing. In *Beyond Parameters in Interdisciplinary Research : Festschrift in Honor of Professor Pranab K. Sen*, 211-230, IMS, Beachwood, OH.
- [9] CARLIN, B.P. AND LOUIS, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman & Hall, London.
- [10] CHOW, SHEIN CHUNG AND LIU, JEN-PEI. (2008). *Design and Analysis of bioavailability and Bioequivalence Studies*. Chapman and Hall, Biostatistics Series.
- [11] COHEN, J.E. (1990). DNA fingerprinting for forensic identification : Potential effects on data interpretation of subpopulation heterogeneity and band number variability.

American Journal of Human Genetics, **46**, 358-368.

- [12] DATTA, G.S. AND GHOSH, M. (1995a). Some remarks on non-informative priors. *J. Amer. Statist. Assoc.*, **90**, 1357-1363.
- [13] DATTA, G.S. AND GHOSH, M. (1995b). Hierarchical Bayes estimators of the error variance in one-way ANOVA models. *J. Statist. Plann. Inf.*, **45**, 399-411.
- [14] DATTA G.S., GHOSH, M. AND MUKHERJEE, R. (2000). Some new results on probability matching priors. *Calcutta Statistical Assoc. Bull.*, **50**, 179-192.
- [15] DATTA, GAURI SANKAR AND MUKHERJEE, RAHUL (2004). *Probability Matching Priors : Higher Order Asymptotics*, Springer.
- [16] DATTA, GAURI SANKAR AND SWEETING, TREVOR J. (2005). Probability Matching Priors. Research Report No. 252, Dept. of Statistical Science, University College London.
- [17] DAWID, A.P. (1991). Fisherian inference in likelihood and prequential frames of reference (with discussion). *J. Roy. Statist. Soc. B.*, **53**, 79-109.
- [18] DEMPSTER, A.P., LAIRD, N.M. AND RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. (Ser. B)*, **39**, 1-38.
- [19] DEVLIN, B. AND RISCH, N. (1992a). A note on Hardy - Weinberg equilibrium of VNTR data using the FBI's fixed bin Method. *American Journal of Human Genetics*, **51**, 549-553.
- [20] DEVLIN, B. AND RISCH, N. (1992b). Ethnic differentiation at VNTR loci, with special reference to forensic applications. *American Journal of Human Genetics*, **51**, 534-548.
- [21] DEVLIN, B. AND RISCH, N. (1993). Physical properties of VNTR data and its impact on tests of allelic independence. *American Journal of Human Genetics*, **53**, 324-329.
- [22] DEVLIN, B., RISCH, N. AND ROEDER, K. (1991). Estimation of allele frequencies

- for VNTR loci. *American Journal of Human Genetics*, **48**, 662-676.
- [23] DEVLIN, B., RISCH, N. AND ROEDER, K. (1992). Forensic inference from DNA fingerprints. *J. Amer. Statist. Assoc.*, **87**, 337-349.
 - [24] DEVLIN, B., RISCH, N. AND ROEDER, K. (1993). NRC Report on DNA typing. *Science*, **260**, 1057-1059.
 - [25] DiCICCIO, T.J. AND STERN, S.E. (1993). On Bartlett adjustments for approximate Bayesian inference. *Biometrika*, **80**, 731-740.
 - [26] DiCICCIO, T.J. AND STERN, S.E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J. Roy. Statist. Soc. B*, **56**, 397-408.
 - [27] DUMOUCHEL, WILLIAM H. (1992). Introduction to Edwards, Lindman and Savage (1963). Bayesian Statistical Inference for Psychological Research. Breakthroughs in Statistics, edited by S. Kotz and N.L. Johnson, Springer Series in Statistics, New York.
 - [28] EDWARDS, WARD, LINDMAN, HAROLD AND SAVAGE, LEONARD J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychological Review*, **70** (3).
 - [29] EFRON, BRADLEY (2013). Large-Scale Inference : Empirical Bayes Methods for Estimation, Testing and Prediction. Cambridge University Press.
 - [30] EFRON, B. AND MORRIS, C. (1971). Limiting the risk of Bayes and Empirical Bayes estimators - Part 1 : The Bayes case. *J. Amer. Statist. Assoc.* **66**, 807-815.
 - [31] — — — (1972). Limiting the risk of Bayes and empirical Bayes estimators - Part II : The empirical Bayes case. *J. Amer. Statist. Assoc.*, **67**, 130-139.
 - [32] — — — (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. *J. Amer. Statist. Assoc.*, **68**, 117-130.
 - [33] — — — (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.*, **70**, 311-319.

- [34] EVETT, I.W., CAGE, P.E. AND AITKEN, C.G.G. (1987). Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases. *J. Roy. Statist. Soc. Ser. C*, **36**, 174-180.
- [35] EVETT, I.W., WERRETT, D.J., GILL, P. AND BUCKLETON, J.S. (1989). DNA fingerprinting on trial. *Nature*, **340**, 435.
- [36] FERGUSON, T.S. (1983). Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics* (H. Rizvi, J. Rustagi and D.O. Siegmund, eds.) 287-302. Academic, New York.
- [37] FORCINA, ANTONIO (1982). Gini's contributions to the Theory of Inference. *International Statistical Review*, **50**, 65-70.
- [38] GARVAN, C.W. AND GHOSH, M. (1997). Noninformative priors for dispersion models. *Biometrika*, **84**, 976-982.
- [39] GEISSER, S. (1990). Some remarks on DNA fingerprinting. *Chance*, **3**, 8-9.
- [40] GEISSER, S. (1992). Some statistical issues in medicine and forensics. *J. Amer. Statist. Assoc.*, **87**, 607-614.
- [41] GEISSER, S. AND JOHNSON, W. (1992). Testing Hardy-Weinberg equilibrium on allelic data from VNTR loci *American Journal of Human Genetics*, **51**, 1084-1088.
- [42] GELFAND, A.E. AND SMITH, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398-409.
- [43] GHOSH, J.K. (1994). *Higher Order Asymptotics*, Institute of Mathematical Statistics and American Statistical Association, Hayward, California.
- [44] GHOSH, J.K., DELAMPADY, M. AND SAMANTA, T. (2006). *An Introduction to Bayesian Analysis, Theory and Methods*, Springer.
- [45] GHOSH, J.K. AND MUKERJEE, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *J. Mult. Anal.*, **38**, 385-393.

- [46] GHOSH, J.K. AND MUKERJEE, R. (1992a). Non-informative priors (with discussion). In *Bayesian Statistics 4*, Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 195-210, Oxford University Press, London.
- [47] GHOSH, J.K. AND MUKERJEE, R. (1992b). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests. *J. Roy. Statist. Soc. B.*, **54**, 867-875.
- [48] GHOSH, J.K. AND MUKERJEE, R. (1993a). On priors that match posterior and frequentist distribution functions. *Can. J. Statist.*, **21**, 89-96.
- [49] GHOSH, J.K. AND MUKERJEE, R. (1993b). Frequentist validity of highest posterior density regions in multiparameter case. *Ann. Inst. Statist. Math.*, **45**, 293-302.
- [50] GHOSH, J.K. AND MUKERJEE, R. (1994a). Higher order comparison of tests : A Bayesian route. In *Essays on Probability and Statistics in Honor of Professor A.K. Bhattacharyya*, Eds. S.P. Mukherjee, A. Chaudhuri and S.K. Basu, 92-101, Presidency College, Calcutta.
- [51] GHOSH, J.K. AND MUKERJEE, R. (1994b). Adjusted versus conditional likelihood : Power properties and Bartlett-type adjustment. *J. Roy. Statist. Soc. B.*, **56**, 185-188.
- [52] GHOSH, J.K. AND MUKERJEE, R. (1995a). Frequentist validity of highest posterior density regions in the presence of nuisance parameters. *Statist. Decisions*, **13**, 131-139.
- [53] GHOSH, J.K. AND MUKERJEE, R. (1995b). On perturbed ellipsoidal and highest posterior density regions with approximate frequentist validity. *J. Roy. Statist. Soc. B.*, **57**, 761-769.
- [54] GHOSH, M., CARLIN, B.P. AND SRIVASTAVA, M.S. (1995). Probability matching priors for linear calibration. *Test*, **4**, 333-357.
- [55] GHOSH, M. AND MUKERJEE, R. (1997). Recent developments on probability matching priors. In *Applied Statistical Science III*, Eds. S.E. Ahmed, M. Ahsanullah and B.K. Sinha, 227-252, Nova Science Publishers, New York.
- [56] GHOSH, M. AND YANG, M.C. (1996). Noninformative priors for the two-sample normal problem. *Test*, **5**, 145-157.

- [57] GINI, C. (1911). Considerazioni sulla probabilita a posteriori e applicazioni al rapporto dei sessi nelle nascite umane. Reprinted in *Metron*, **14**, 173-240.
- [58] GOOD, I.J. (1992). Introduction to Robbins (1956). An Empirical Bayes Approach to Statistics. Breakthroughs in Statistics, Vol. 1. 379-387, Kotz, Samuel and Johnson, Norman L. (Eds.) Springer series in Statistics.
- [59] JAMES, WILLARD AND STEIN, CHARLES (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, **1**, 361-379.
- [60] JEFFREYS, A.J., WILSON, V. AND THEIN, S.L. (1985a). Hypervariable ‘minisatellite’ regions in human DNA. *Nature*, **314**, 66-73.
- [61] JEFFREYS, A.J., WILSON, V. AND THEIN, S.L. (1985b). Individual-specific ‘fingerprints’ of human DNA. *Nature*, **316**, 76-79.
- [62] KASS, R.E. AND WASSERMAN, L. (1996). The selection of prior distributions by formal rules. *J. Amer. Statist. Assoc.*, **91**, 1343-1370.
- [63] LEHMANN, E.L. (1986). *Testing Statistical Hypotheses*, 2nd. ed. John Wiley & Sons, New York.
- [64] LINDLEY, D.V. (1977). A problem in forensic science. *Biometrika*, **64**, 2 : 207-13.
- [65] MARITZ, J.S. AND LWIN, T. (1989). *Empirical Bayes Methods*, 2nd ed., Chapman and Hall.
- [66] MASRY, E. AND RICE, J.A. (1992). Gaussian Deconvolution via differentiation, *Canadian Journal of Statistics*, **20** : 9-21.
- [67] MORRIS, C. (1977). Interval Estimation for Empirical Bayes Generalizations of Stein’s Estimator. *Proceedings of the Twenty-Second Conference on the Design of Experiments in Army Research Development and Testing, ARO Report*, 77-2.
- [68] MORRIS, C. (1981). Parametric Empirical Bayes Confidence Intervals. *Proceedings of the Conference on Scientific Inference, Data Analysis and Robustness*, University of Wisconsin.

- [69] MORRIS, C.N. (1983). Parametric Empirical Bayes Inference : Theory and Applications, *Journal of the American Statistical Association*, **78**, 47-65.
- [70] MUKERJEE, R. AND DEY, D.K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter : Higher order asymptotics. *Biometrika*, **80**, 499-505.
- [71] MUKERJEE, R. AND GHOSH, M. (1997). Second-order probability matching priors. *Biometrika*, **84**, 970-975.
- [72] MUKERJEE, R. AND REID, N. (2000). On the Bayesian approach for frequentist computations. *Brazilian J. Probab. Statist.*, **14**, 159-166.
- [73] NEYMAN, J. (1962). Two breakthroughs in the theory of statistical decision making. *Rev. Internat. Statist. Inst.*, **30**, 11-27.
- [74] NICOLAOU, A. (1993). Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. Roy. Statist. Soc. B.*, **55**, 377-390.
- [75] PEERS, H.W. (1965). On confidence sets and Bayesian probability points in the case of several parameters. *J. Roy. Statist. Soc. B.*, **27**, 9-16.
- [76] RAO, A. RAMACHANDRA AND BHIMASANKARAM, P. (2000). *Linear Algebra*, 2nd ed. Hindustan Book Agency (India).
- [77] RAO, C. RADHAKRISHNA (1973). *Linear Statistical Inference and its Applications*, 2nd ed. Wiley Eastern Limited.
- [78] REID, N. (1995). Likelihood and Bayesian approximation methods (with discussion). In *Bayesian Statistics 5*, Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 351-368, Oxford University Press, London.
- [79] RISCH, N. AND DEVLIN, B. (1992a). On the probability of matching DNA fingerprints. *Science*, **255**, 717-720.
- [80] RISCH, N. AND DEVLIN, B. (1992b). DNA fingerprint matches. *Science*, **256**, 1744-1746.

- [81] ROBBINS, H.E. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab. 1*, 157-164. Univ. California Press, Berkeley.
- [82] ROBBINS, HERBERT (1983). Some thoughts on Empirical Bayes Estimation. *The Annals of Statistics*, Vol. **11**, No. 3, 713-723.
- [83] ROEDER, KATHRYN (1994). DNA Fingerprinting : A Review of the Controversy, *Statistical Science*, **9**, No. 2, 222-278.
- [84] ROEDER, K., ESCOBAR, M., KADANE, J. AND BALAZS, I. (1993). Measuring heterogeneity in forensic databases : a hierarchical approach to the question. Technical Report, Carnegie Mellon Univ.
- [85] SCOTT, J.G. AND BERGER, J.O. (2010). Bayes and Empirical - Bayes multiplicity adjustment in the variable - selection problem, *Ann. Statist.* **38** No. 5, 2587-2619.
- [86] SERFLING, ROBERT J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- [87] SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.
- [88] STEIN, CHARLES (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. on Math. Statist. and Prob.*, Vol. **1**, 197-206.
- [89] STEIN, C. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics, Banach Center Publications*, **16**, 485-514, Polish Scientific Publishers, Warsaw.
- [90] SUN, D. AND YE, K. (1996). Frequentist validity of posterior quantiles for a two-parameter exponential family. *Biometrika*, **83**, 55-65.
- [91] TIBSHIRANI, R.J. (1989). Noninformative priors for one parameter of many. *Biometrika*, **76**, 604-608.
- [92] WELCH, B.L. AND PEERS, H.W. (1963). On formulae for confidence points based on integrals of weighted likelihoods. *J. Roy. Statist. Soc. B.*, **25**, 318-329.

Appendix

The proofs of Theorems 2.1 and 4.1 of Chapter 3 are presented here. Equation numbers refer to equations / expressions in Chapter 3.

The following preliminaries will facilitate the presentation of the proofs. Let

$$\delta = (V + A)^{1/2}, \quad X_1 = k^{1/2}(z'_i b - z'_i \beta) / \delta, \quad X_2 = (Y_i - z'_i \beta) / \delta, \quad (A.1)$$

$$X_3 = k^{1/2}(\delta^{-2} s^2 - 1), \quad X_4 = X_3 - k^{-1/2}(X_3^2 - 2). \quad (A.2)$$

By (2.1), $c_{ik}^{-1/2} X_1$ is standard normal and hence the assumption made in Section 3.1 implies that X_1 is stochastically bounded. It is easy to see that so are X_2, X_3 and X_4 ; cf. (2.1).

Lemma. (i) $E_\xi(X_1) = 0$, (ii) $E_\xi(X_4) = o(k^{-1})$, (iii) $E_\xi(X_1^2) = c_{ik}$, (iv) $E_\xi(X_2^2) = 1$, (v) $E_\xi(X_4^2) = 2 + o(1)$, (vi) $E_\xi(X_1 X_4) = 0$, (vii) $E_\xi(X_2 X_4) = 0$, (viii) $E_\xi(X_2 X_4^2) = 0$, (ix) $E_\xi(X_1 X_2 X_4) = o(1)$, (x) $E_\xi(X_2^2 X_4^2) = 2 + o(1)$.

Proof. The proofs of parts (i) - (v) are either obvious or straightforward. To prove the remaining parts, let us define

$$U = (U_1, \dots, U_k)' = \delta^{-1}(Y - Z_k \beta), \quad w_i = Z_k(Z'_k Z_k)^{-1} z_i$$

and let M be the orthogonal projector on the orthocomplement of the column space of Z_k . Then

$$\text{tr}(M) = k - r, \quad M w_i = 0 \quad (A.3)$$

and by (2.2), (A.1) and (A.2)

$$X_1 = k^{1/2} w'_i U, \quad X_2 = U_i, \quad X_3 = k^{1/2} \{(k - r)^{-1} U' M U - 1\}. \quad (A.4)$$

By (A.2) and (A.4), each of $X_1 X_4, X_2 X_4$ and $X_2 X_4^2$ is an odd polynomial in U . Hence parts (vi) - (viii) follow noting that the elements of U are independently standard normal, a fact that is also used in the rest of the proof whenever necessary.

Next, by (A.2) and (A.4)

$$E_\xi(X_1 X_2 X_4) = k E_\xi \left[(w'_i U) U_i \{(k - r)^{-1} U' M U - 1\} \right] + o(1). \quad (A.5)$$

Write w_{ii} for the i th element of w_i and m_i for the i th column of M . Then by (A.3),

$$E_\xi\{(w'_i U)U_i(U' MU)\} = w_{ii}tr(M) + 2w'_i m_i = w_{ii}(k - r).$$

Since $E_\xi\{(w'_i U)U_i\} = w_{ii}$, part (ix) is now evident from (A.5).

Finally, in order to prove (x), we note that by (A.2) and (A.4)

$$E_\xi(X_2^2 X_4^2) = kE_\xi \left[U_i^2 \{(k - r)^{-1} U' MU - 1\}^2 \right] + o(1). \quad (A.6)$$

Now, M is idempotent with (i, i) th element $1 - z'_i(Z'_k Z_k)^{-1} z_i = (1 - k^{-1} c_{ik})$. Hence after some algebraic manipulation,

$$E_\xi\{U_i^2(U' MU)\} = k - r + 2(1 - k^{-1} c_{ik}),$$

$$E_\xi\{U_i^2(U' MU)^2\} = (k - r)^2 + 2(k - r) + 4(k - r + 2)(1 - k^{-1} c_{ik}).$$

Using the above in (A.6) and recalling the assumption on the boundedness of $\{c_{ik}\}$, part (x) of the lemma follows.

We are now in a position to present the proofs. All stochastic expansions considered below are on a set with P_ξ -probability $1 + o(k^{-1})$ (uniformly over compact ξ -sets). In particular, $s^2 > V$ on this set, so that by (2.2), (A.1) and (A.2),

$$\hat{B} = \frac{k - r - 2}{k - r} (1 + k^{-1/2} X_3)^{-1} B = B(1 - k^{-1/2} X_4) + o(k^{-1}). \quad (A.7)$$

Proof of Theorem 2.1. Given Y , conditionally θ_i is $N((1 - B)Y_i + Bz'_i\beta, V(1 - B))$. Hence

$$P_\xi\{(\theta_i - \hat{\theta}_i)/s_i \leq t_k\} = E_\xi\{\Phi(T)\}, \quad (A.8)$$

where

$$T = \{\hat{\theta}_i + t_k s_i - (1 - B)Y_i - Bz'_i\beta\} / \{V(1 - B)\}^{1/2}. \quad (A.9)$$

Recalling the definitions of B, δ and ρ , by (2.5), (2.6), (A.1) and (A.7), after some algebra,

$$\hat{\theta}_i - (1 - B)Y_i - Bz'_i\beta = \delta B\{k^{-1/2}(X_1 + X_2 X_4) - k^{-1} X_1 X_4\} + o(k^{-1}), \quad (A.10)$$

$$\begin{aligned} s_i &= \{V(1 - B)\}^{1/2} \left[1 + \frac{1}{2} k^{-1/2} \rho X_4 + \frac{1}{2} k^{-1} \{(r + 2X_2^2)\rho \right. \\ &\quad \left. - \frac{1}{4} \rho^2 X_4^2\} \right] + o(k^{-1}). \end{aligned} \quad (A.11)$$

By (A.9) - (A.11),

$$T = t_k + k^{-1/2}g_1 + k^{-1}g_2 + o(k^{-1}), \quad (\text{A.12})$$

where

$$g_1 = \frac{1}{2}t_k\rho X_4 + \rho^{1/2}(X_1 + X_2X_4), \quad (\text{A.13})$$

$$g_2 = \frac{1}{2}t_k\{(r + 2X_2^2)\rho - \frac{1}{4}\rho^2X_4^2\} - \rho^{1/2}X_1X_4. \quad (\text{A.14})$$

By (A.12),

$$\Phi(T) = \Phi(t_k) + k^{-1/2}\phi(t_k)g_1 + k^{-1}\phi(t_k)(g_2 - \frac{1}{2}t_k g_1^2) + o(k^{-1}). \quad (\text{A.15})$$

Using the lemma, from (A.13) and (A.14),

$$E_\xi(g_1) = o(k^{-1/2}), \quad E_\xi(g_2) = \frac{1}{2}t_k\{(r + 2)\rho - \frac{1}{2}\rho^2\} + o(1), \quad (\text{A.16})$$

$$E_\xi(g_1^2) = \frac{1}{2}t_k^2\rho^2 + (c_{ik} + 2)\rho + o(1). \quad (\text{A.17})$$

If one substitutes (A.15) in (A.8) and then employs (A.16) and (A.17) then the result follows.

Proof of Theorem 4.1. By (2.3), (2.8) and (A.7), the length L of the interval I satisfies $L = 2z\{1 + k^{-1}h_{ik}(z; \rho)\}s_i + o(k^{-1})$. Hence by (A.11),

$$\begin{aligned} L = 2z\{V(1 - B)\}^{1/2} & \left[1 + \frac{1}{2}k^{-1/2}\rho X_4 + k^{-1} \left\{ \left(\frac{1}{2}r + X_2^2 \right) \rho - \frac{1}{8}\rho^2 X_4^2 \right. \right. \\ & \left. \left. + h_{ik}(z; \rho) \right\} \right] + o(k^{-1}). \end{aligned}$$

Part (a) now follows using (2.4) and the lemma. The proof of (b) is similar.